43357

National Library of Canada
Canadian Theses Division
Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada
Division des thèses canadiennes

PERMISSION TO MICROFILM — AUTORISATION DE MICROFILMER

• Please print or type — Écrire en lettres moulées ou dactylographier

Full Name of Author — Nom complet de l'auteur

Gary David Brayer

Date of Birth — Date de naissance	Country of Birth — Lieu de naissance
April 1 1953	Canada

Permanent Address — Résidence fixe

4820 - Malmo Road
Edmonton, Alberta
Canada

Title of Thesis — Titre de la thèse

Structural Studies of the Microbial Serine Proteases, Streptomyces griseus Protease A, Streptomyces griseus Protease B and Alpha Lytic Protease

University — Université

University of Alberta

Degree for which thesis was presented — Grade pour lequel cette thèse fut présentée

Ph. D

Year this degree conferred — Année d'obtention de ce grade	Name of Supervisor — Nom du directeur de thèse
1979	M. N. G. James

Permission is hereby granted to the NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈQUE NATIONALE DU CANADA de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans l'autorisation écrite de l'auteur.

Date	Signature
September 20, 1979	G. Brayer

NL-91 (4/77)

National Library of Canada

Cataloguing Branch
Canadian Theses Division

Ottawa, Canada
K1A 0N4

Bibliothèque nationale du Canada

Direction du catalogage
Division des thèses canadiennes

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE

THE UNIVERSITY OF ALBERTA

STRUCTURAL STUDIES OF THE MICROBIAL SERINE PROTEASES, *STREPTOMYCES GRISEUS* PROTEASE A, *STREPTOMYCES GRISEUS* PROTEASE B AND ALPHA LYTIC PROTEASE

by

© GARY DAVID BRAYER

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

IN

BIOCHEMISTRY

DEPARTMENT OF BIOCHEMISTRY

EDMONTON, ALBERTA

1979

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled STRUCTURAL STUDIES OF THE MICROBIAL SERINE PROTEASES, STREPTOMYCES GRISEUS PROTEASE A, STREPTOMYCES GRISEUS PROTEASE B AND ALPHA LYTIC PROTEASE submitted by GARY DAVID BRAYER in partial fulfilment of the requirements for the degree of DOCTOR OF PHILOSOPHY in BIOCHEMISTRY.

Michael James
.....
Supervisor

.....
.....
.....
.....
External Examiner

Date. September 6, 1979

Abstract

The 2.8 angstrom resolution structures of three microbial serine proteases, Streptomyces griseus Protease A, Streptomyces griseus Protease B and alpha lytic protease, have been determined. These enzymes are shown to be structurally related to the pancreatic family of serine proteases, rather than the previously defined bacterial subtilisin family. All three microbial enzymes are structurally similar and appear to be representative of evolutionary precursors of the mammalian pancreatic serine proteases.

The determination of these structures has allowed for the detailed structural comparison of the microbial enzymes with their pancreatic counterparts. It is found that the dispositions of the active site residues Ser-214, Asp-102, His-57 and Ser-195 are nearly identical in all these enzymes. Despite the presence of only very low overall primary sequence homology (max. 21%), it is shown that approximately 60% of the residues of the microbial enzymes are topologically equivalent to residues in the pancreatic enzymes. Earlier primary sequence alignments of these enzymes were hampered by the presence of low sequence homology. A primary sequence alignment based on topological equivalence is presented.

Major structural differences between the microbial and pancreatic enzymes reside in two regions. The first of these is related to the presence of a zymogen activation mechanism

in the pancreatic enzymes and the apparent absence of this control mechanism in the microbial enzymes. Also observed, are significant rearrangements of polypeptide loops in the active site region of the microbial enzymes. These alterations serve to explain the unique substrate binding properties of these enzymes.

Structural analyses of SGPA/peptide aldehyde complexes have also been completed. One of the peptide aldehydes investigated, chymostatin, is a naturally occurring bacterial inhibitor of serine proteases. These results show that peptide aldehydes form covalent tetrahedral hemiacetal adducts with Ser-195. The complexes formed are similar to covalent tetrahedral transition state intermediates postulated to occur during peptide catalysis. From these studies it was also possible to determine the positions of several binding subsites, which earlier investigators have shown play an important role in substrate binding and catalysis.

Two SGPB/chloromethyl ketone peptide complexes were also subjects of structural analysis. These studies demonstrated that chloromethyl ketone inhibitors form two covalent bonds in the active site of SGPB. One of these is from the methylene carbon atom of the inhibitor to the imidazole ring of His-57. The other is formed from the terminal carbonyl carbon atom to the side chain of Ser-195. Comparison of inhibitor binding modes to the microbial and pancreatic serine proteases is also discussed.

Acknowledgements

Throughout the course of this research, I have always been particularly indebted to my wife, Wendy, whose encouragement, love and patience has made this work a reality.

It also gives me great pleasure to acknowledge the contributions that my collegues Louis Delbaere, Koto Hayakawa, I-N. Hsu, Mike James, Masaaki Matsushima, and Anita Sielecki have made to the experiments discussed in the following text. Only with their assistance, interest and helpful discussions has this work been successful. I am also indebted to Mike James for his guidance and support during this research and a careful reading of this thesis.

I would also wish to express my sincere thanks to the other members of the Department of Biochemistry for creating an enjoyable and stimulating working environment.

The financial support of the Medical Research Council of Canada is gratefully acknowledged.

Table of Contents

Abstract.....iv
Acknowledgements.....vi
Table of Contents.....vii
List of Tables.....x
List of Figures.....xiv
List of Abbreviations.....xxiv

Chapter — Page

I. Introduction.....1

A. The Serine Proteases.....1

B. The Extracellular Serine Proteases Of *Streptomyces griseus* (Strain K1).....18

C. The Extracellular Serine Protease Of *Myxobacter 495*.....29

D. Objectives Of This Research Project.....34

II. Amino Acid Sequence Alignment Of The Bacterial and Pancreatic Serine Proteases.....38

III. Molecular Structure Of *Streptomyces griseus* Protease A At 2.8 Angstrom Resolution.....48

A. Isolation And Crystallization.....48

B. Heavy-atom Derivative Preparation.....49

C. Data Collection.....51

D. Background Correction And Data Reduction.....52

E. Data Scaling.....57

F. Heavy-Atom Solution, Phase Calculation And Least-squares Refinement.....60

G. Phasing Results.....65

H. Heavy-Atom Binding Sites.....70

I. Electron Density Maps And Model Building.....72

J. Molecular Conformation Of SGPA.....76

K. Structural Comparison Of SGPA and Alpha-Chymotrypsin.....84

L. SGPA As An Evolutionary Precursor Of Alpha-Chymotrypsin.....99

M. Active Site Region.....101

N. Conformation Of Ser-195.....108

IV. The Tertiary Structure Of Streptomyces griseus Protease B At 2.8 Angstrom Resolution.....110

A. Structure Determination.....110

B. Interpretation Of The Electron Density Map....110

C. Molecular Conformation And Comparison With SGPA And Alpha-Chymotrypsin.....112

D. Active Site Conformation Of SGPB.....117

V. The Tertiary Structure Of Alpha Lytic Protease At 2.8 Angstrom Resolution.....122

A. Isolation and Crystallization.....122

B. Data Collection.....124

C. Heavy-Atom Derivatives.....124

D. Data Reduction And Scaling.....128

E. Phase Angle Determination.....129

F. Heavy-Atom Binding Sites.....136

G. Native Electron Density Map And Interpretation.....138

H. Atomic Model Of Alpha Lytic Protease.....140

I. Structural Comparison Of Alpha Lytic Protease With SGPA And Other Serine Proteases.....147

J. Active Site Conformation.....159

K. Substrate Recognition Sites.....164

L. Environment of Asp-102 And Interpretation Of NMR Data.....168

VI. Complexation Of A Tetrapeptide Aldehyde In The Active Site Of SGPA.....175

A. Peptide Aldehyde Substrate Analogs.....175

B. Crystallographic Data Collection.....177

C. Difference Map Interpretation.....179

D. Comparison With Solution Data And Other Serine Proteases.....186

E. Peptide Aldehydes As Transistion State Analogs.....189

VII. The Complex Formed By Chymostatin And SGPA.....194

A. Chymostatin: Chemical Structure And Specificity.....194

B. Collection Of Three-Dimensional Data.....196

C. Difference Electron Density Interpretation....198

D. Comparison With The Tetrapeptide Aldehyde - SGPA Complex.....201

VIII. Chloromethyl Ketone Inhibitor Studies Of SGPB....205

A. Chloromethyl Ketone Peptide Analogs.....205

B. Crystallization and Data Collection.....207

C. Interpretation of Difference Electron Density Maps.....210

The GLF/SGPB Complex.....212

The AGF/SGPB Complex.....216

D. Chloromethyl Ketone Peptides As Substrate Analogs.....218

References.....226

Appendix 1.....239

Appendix 2.....240

Appendix 3.....242

Appendix 4.....244

Appendix 5.....246

List of Tables

Chapter II - Amino Acid Sequence Alignment Of The Bacterial and Pancreatic Serine Proteases

Table — Page

1. Amino acid sequence alignment of *Streptomyces griseus* Protease A, *Streptomyces griseus* Protease B, alpha lytic protease, alpha-chymotrypsin, elastase and trypsin.....43

2. Amino acid sequence identity and topological equivalence matrix for the proteins of Table 1.....46

Chapter III - Molecular Structure Of Streptomyces griseus Protease A At 2.8 Angstrom Resolution

Table — Page

3. Crystal data for SGPA.....51

4. Cell dimension changes in derivative crystals...53

5. Statistics of data collection to 2.8 angstrom resolution for SGPA.....54

6. Application of the background function.....56

7. Data set scaling and heavy-atom differences.....59

8. Refined heavy-atom parameters for SGPA.....66

9. Phase determination statistics for SGPA.....69

10. Sites of heavy-atom binding.....71

11. Main chain to side chain hydrogen bonds.....82

12. Side chain to side chain hydrogen bonds.....82

13. Beta bends found in SGPA.....83

Chapter V - The Tertiary Structure Of Alpha Lytic Protease At 2.8 Angstrom Resolution

Table Page

14. Crystal data for alpha lytic protease.....125

15. Data collection methods 126

16. Statistics of data collection to 2.8 angstrom resolution for alpha lytic protease...127

17. Cell dimension changes in derivative crystals..128

18. Data set scaling and heavy-atom differences....130

19. Refined heavy-atom parameters for alpha lytic protease derivatives.....132

20. Phase determination statistics for alpha lytic protease.....135

21. Major sites of heavy-atom binding.....136

22. Minor sites of heavy-atom binding.....137

23. Main chain to side chain hydrogen bonds.....147

24. Side chain to side chain hydrogen bonds and salt bridges.....147

25. Hairpin loops found in alpha lytic protease....148

Chapter VI - Complexation Of A Tetrapeptide Aldehyde In The Active Site Of SGPA

Table Page

26. SGPA/peptide aldehyde complex diffraction statistics.....178

Chapter VII - The Complex Formed By Chymostatin and SGPA

Table Page

27. Crystal data for the SGPA/cymostatin complex...197

Chapter VIII - Chloromethyl Ketone Inhibitor Studies Of SGPB

Table Page

28. SGPB: native and inhibitor complex diffraction statistics.....208

List of Figures

Chapter I - Introduction

Figure — Page

1. Stereo-drawing of the three-dimensional conformation of Asp-102, His-57 and Ser-195 in the active site of porcine pancreatic elastase.....13

2. A schematic representation of the proposed catalytic mechanism of serine proteases.....14

Chapter III - Molecular Structure Of *Streptomyces griseus* Protease A At 2.8 Angstrom Resolution

Figure — Page

3. Photomicrograph of crystalline *Streptomyces griseus* protease A.....49

4. Precession photograph of a native SGPA crystal.....50

5. Heavy-atom Patterson difference maps for the three major derivatives used in this study.....62

6. The variation of the ratio r.m.s. f(H) to r.m.s. E(H) and the figure of merit (m) as functions of {sin(theta)/lambda}^{2}.....67

7. Distribution of the figures of merit among native enzyme reflections phased using the multiple isomorphous technique.....68

8. Stereo-drawing of the rhenium binding site coordination sphere.....73

9. Plot of the phi, psi torsional angles for the atomic model of SGPA at 2.8 angstrom resolution.....78

10. Stereo-drawing of the polypeptide chain of SGPA in the region of the cis-Pro-99A peptide bond.....79

11. Schematic drawing of the observed secondary structural features of SGPA.....80

12. Stereo-drawing of all the main chain atoms of SGPA.....81

13. Stereo-drawing of the C-terminal helical region of the SGPA molecule.....84

14. Alpha-carbon backbone drawings of SGPA and alpha-chymotrypsin.....85

15. An alpha-carbon backbone drawing of SGPA superimposed on that of alpha-chymotrypsin to show the topological equivalence between the two enzyme molecules.....86

16. A comparison of the environments of Asp-194 in SGPA and alpha-chymotrypsin.....89

17. The environment of the N-terminus of SGPA.....90

18. A stereo-representation of the multiple isomorphous replacement phased 2.8 angstrom resolution electron density map of SGPA through the region of the molecule containing the active site.....102

19. Stereo-drawing of the active site region of SGPA and alpha-chymotrypsin in a similar orientation.....103

Chapter IV - The Tertiary Structure Of Streptomyces griseus Protease B At 2.8 Angstrom Resolution

Figure

Page

20. Diagrammatic representation of the two six stranded beta barrels of SGPB.....113

21. a) Stereo-drawing of the polypeptide chain folding of SGPB.
b) A stereo-view illustrating the topological comparison of the polypeptide chains of SGPB and SGPA.....115

22. Topological comparison of the polypeptide folding of SGPB and alpha-chymotrypsin.....116

23. Stereo-view of eight sections of the electron density map of SGPB in the region of the active site.....118

24. Stereo-view of the active site region of SGPB..120

Chapter V - The Tertiary Structure Of Alpha Lytic Protease At 2.8 Angstrom Resolution

Figure — Page

25. Photomicrograph of crystalline alpha lytic protease.....123

26. Precession photograph of a native alpha lytic protease crystal.....124

27. The variation of the ratio r.m.s. f(H) to r.m.s. E(H) and the figure of merit (m), as functions of $\{\sin(\text{theta})/\text{lambda}\}^2$.....133

28. The distribution of figures of merit among native enzyme reflections of alpha lytic protease.....134

29. a) A stereo-drawing of the alpha-carbon backbone of alpha lytic protease.
    b) An alpha-carbon drawing of alpha lytic protease superimposed on that of SGPA in a manner designed to maximize topological equivalence.....141

30. Plot of the phi, psi torsional angles for the atomic model of alpha lytic protease.....143

31. A stereo-drawing of the Phe-94 to Pro-99A cis-peptide link at the beta bend extremity of the aspartate loop in alpha lytic protease..144

32. Schematic drawing of the observed secondary structural features of alpha lytic protease....145

33. a) A stereo-drawing of the alpha-carbon backbone of alpha lytic protease superimposed on the alpha-carbon backbone of SGPB.

b) In this drawing the alpha-carbon backbone of alpha lytic protease has been superimposed on the alpha-carbon backbone of elastase.....150

34. Stereo-drawing of the environment about the internal salt bridge formed by the guanidinium group of Arg-138 to the active site residue Asp-194 in alpha lytic protease.....152

35. Stereo-drawing of the environment about the N-terminus of alpha lytic protease.....154

36. The tertiary structure of the double beta bend about Ser-195, involving residues Gly-192A to Gly-197, is shown.....157

37. A stereo-representation of the multiple isomorphous replacement phased 2.8 angstrom resolution native electron density map of alpha lytic protease through the active site region of the molecule.....160

38. A stereo-drawing showing the bound conformation of a sulfate anion in the active site of alpha lytic protease.....161

39. Stereo-drawing of the active site of alpha lytic protease.....162

40. Stereo-drawing of the active site of elastase..168

Chapter VI - Complexation Of A Tetrapeptide Aldehyde In The Active Site Of SGPA

Figure Page

41. A stereo-representation of the difference electron density map of the SGPA/peptide

aldehyde complex at 2.8 angstrom resolution....180

42. Stereo-representation of a portion of the SGPA/peptide aldehyde electron density map in the region of the covalent bond between the active site residue Ser-195, and the bound inhibitor.....181

43. Stereo-drawing of the SGPA/peptide aldehyde complex as determined from the 2.8 angstrom resolution difference electron density map.....183

44. Stylized drawing of the interactions formed between the bound peptide aldehyde and SGPA....191

Chapter VII - The Complex Formed By Chymostatin and SGPA

Figure Page

45. A structural comparison of chymostatin A and the synthetic peptide tetrapeptide aldehyde inhibitor used in the previous binding study of SGPA.....195

46. Stereo-representation of the difference electron density map of the chymostatin/SGPA complex at 2.8 angstrom resolution in the

region of the active site of the enzyme.....199

47. Stereo-drawing of the chymostatin/SGPA complex as determined by fitting the difference electron density map.....200

48. Stereo-drawing of chymostatin and the synthetic Ac-Pro-Ala-Pro-Phe-al inhibitor superimposed.....202

Chapter VIII - Chloromethyl Ketone Inhibitor Studies Of SGPB

Figure Page

49. Stereo-representation of the 2.8 angstrom resolution difference electron density maps of the GLF/SGPB and the AGF/SGPB inhibitor complexes.....211

50. Stereo-drawing of the GLF inhibitor bound in the active site of SGPB.....213

51. Stereo-representation of the electron density of the GLF/SGPB complex in the vicinity of the active site residues, His-57 and Ser-195.....215

52. Stereo-drawing of the active site of SGPB showing the conformation of the bound AGF inhibitor.....217

53. Stereo-representation of the electron density map of the AGF/SGPB complex in the active site region.....217

List of Abbreviations

a, b, c	unit cell vectors in angstrom units
Abs.	absolute
Ac	acetyl
-al	aldehyde group
a-LP	alpha lytic protease
alpha(P)	rative structure factor phase angle
alpha(PH-calc)	calculated derivative structure factor phase angle
AGF	Boc-Ala-Gly-Phe-CK
ATEE	N-acetyl-L-tyrosine ethyl ester
BAEE	N-benzoyl-L-arginine ethyl ester
Boc	N-t-Butyloxycarbonyl
BPTI	bovine pancreatic trypsin inhibitor
BT	bovine trypsin
°C	degrees centigrade
calc.	calculated
CHYM	alpha-chymotrypsin
CK	chloromethyl ketone
cm	centimeter
CM	carboxy methylated
C-terminal	carboxy terminal
d	day
DFP	diisopropyl fluorophosphate
e	electron
ELAS	porcine elastase

E(H)	lack of closure error
f(H)	calculated heavy-atom structure factor amplitude
$\underline{f(H)}$	calculated heavy-atom structure factor
F(H)	$\{F(PH)^+ + F(PH)^-\}/2$
F(P)	native enzyme structure factor amplitude
$\underline{F(P)}$	native enzyme structure factor
F(PH)	derivative structure factor amplitude
F(P+I)	inhibitor complex structure factor amplitude
$F(PH)^+$, $F(PH)^-$	Friedel pair of derivative structure factor amplitudes
GLF	Boc-Gly-Leu-Phe-CK
h	hour
Hz	hertz
kV	kilovolt
m	figure of merit
$\bar{m}$	mean figure of merit over a range
<m>	overall figure of merit of all data
M	molar
mA	milliampere
Max.	maximum
MC	mercuric chloranilate
ME	methyl ester
mg	milligram
MIR	multiple isomorphous replacement
mM	millimolar
mm	millimeter
MMS-X	molecular modeling graphics system

No.	number
N-terminal	amino terminal
p-	para
Plat	platinum diamino dichloride
PMA	phenylmercuric acetate
PNPA	p-nitro phenyl acetate
r.m.s.	root mean square
s	seconds
sat.	saturated
scale(D)	absolute scale determined for a heavy-atom derivative data set
scale(N)	absolute scale determined for a native enzyme data set
SGPA	Streptomyces griseus Protease A
SGPB	Streptomyces griseus Protease B
SGT	Streptomyces griseus Trypsin
sp.	species
Tos	tosyl
TPCK	L-(1-tosylamido-2-phenyl) ethyl chloromethyl ketone
TT	two-theta
URE	ureido-group
V	volume of the crystallographic unit cell
Vm	volume per unit molecular weight
w	weight

Note: For amino acid designations see Appendix 1. Individual amino acid atom abbreviations follow the convention of Diamond (1966).

I. Introduction

A. The Serine Proteases

The serine proteases are a class of proteolytic enzymes characterized by the presence of an uniquely reactive serine residue, which takes part in the catalytic event. Enzymes of this type are endoproteases, catalyzing the hydrolysis of peptide bonds in polypeptide chains and protein molecules. Such enzymes are widely distributed in nature, being found in mammals, fish, plants, insects and bacteria (Markland and Smith, 1971; Shaw, 1970). The abnormally high nucleophilicity of the active serine residue apparently arises from specialized structural features in the active sites of these enzymes. This reactive residue is susceptible to derivatization by a number of reagents in a stoichiometric manner leading to complete enzymatic inhibition (Jansen et al., 1949; Shaw, 1970). One such reagent, diisopropyl fluorophosphate, is highly specific for the reactive serine of serine proteases and is routinely used to canvass newly isolated enzymes to determine if they may also be of this type.

Even before the first structural studies of a serine protease had been carried out, ample evidence had accumulated to implicate a histidine side chain in the enzymatic mechanism. The presence of an essential histidine residue was initially suggested by pH inactivation studies. It was found that catalysis was dependent on a group with a

pKa of approximately seven, in the range expected for histidine titration (Bender and Killheffer, 1973). Subsequent studies of irreversible chloromethyl ketone inhibition of serine proteases demonstrated that a single histidine residue is present in the active sites of these enzymes (Schoellmann and Shaw, 1963; Powers, 1977). Unfortunately, inhibitor studies such as those which initially determined the presence of both reactive histidine and serine residues, are in themselves not capable of defining the structural attributes of serine proteases that are responsible for the presence of catalytic activity. Thus, more recent studies of the structure and mechanism of serine proteases have centered around the elucidation of the tertiary structures of these enzymes.

The most intensively studied of the serine proteases have been those isolated from bovine pancreas, in particular alpha-chymotrypsin. Alpha-chymotrypsin catalyzes the hydrolysis of peptide bonds adjacent to the carbonyl groups of the aromatic amino acids phenylalanine, tyrosine and tryptophan (Hess, 1971).[1] This enzyme is synthesized in the acinar cells of the pancreas as a catalytically inert precursor (zymogen), chymotrypsinogen A. Chymotrypsinogen A is carried by the pancreatic juice into the small intestine where it is converted into an active form by a number of proteolytic cleavages. Depending on the extent of proteolytic cleavage during activation, chymotrypsinogen A

---

[1]Unless otherwise indicated, all amino acids discussed in this dissertation are of the L form.

can be converted into pi-, delta-, gamma- or alpha-chymotrypsin (Bender and Killheffer, 1973). Alpha-chymotrypsin has been the most carefully studied in terms of three-dimensional structure of all these forms. This enzyme has a molecular weight of approximately 25,200 and is constructed from 241 amino acids. Three polypeptide chains are generated from the single polypeptide chain of its zymogen precursor (Hess, 1971). Also incorporated into the structure of this enzyme are five disulfide bridges, one of which is to the N-terminal amino acid residue.

The first structural studies of a serine protease using the X-ray crystallographic technique were carried out on the p-toluene sulphonyl inhibited form of alpha-chymotrypsin. These structural studies led to an interpretable, high resolution electron density map of alpha-chymotrypsin (Matthews et al., 1967; Sigler et al., 1968), allowing the elucidation of the detailed three-dimensional structure of this enzyme.

Overall, alpha-chymotrypsin was found to be roughly spherical in shape having a diameter of approximately 40 angstroms. With the construction of a detailed molecular model of the enzyme it became apparent that the majority of polypeptide chain was in the anti-parallel beta sheet conformation. However, three turns of alpha-helical structure are observed near the C-terminal end of the molecule. A further poorly defined helical region is composed of residues 164 to 176 (Birktoft and Blow, 1972).

The internal structure of chymotrypsin was found to be constructed around two hydrophobic cores. Each of these hydrophobic cores is formed from six strands of polypeptide chain hydrogen bonded in an anti-parallel beta sheet conformation and folded so as to approximate a closed cylinder. These closed cylinders have been termed beta-barrels (Birktoft and Blow, 1972). At the juncture of the two beta-barrels of alpha-chymotrypsin and on the surface of the enzyme, are found the active site residues. The reactive serine of this enzyme was easily identified as it was the point of attachment of the bound p-toluene sulphonyl group. Near this serine residue (Ser-195, as numbered by Hartley and Kauffman, 1966) was found the side chain of a histidine residue (His-57), as postulated by earlier inhibition studies[2] (Schoellmann and Shaw, 1963). An additional feature observed in the active site of alpha-chymotrypsin was the presence of a buried aspartate residue (Asp-102) whose side chain interacts with that of His-57. Subsequent structural analysis of gamma-chymotrypsin (chemically identical with alpha-chymotrypsin but having minor pH induced conformational differences) also showed it to have a very similar overall structure and positioning of a serine, histidine and aspartate residue in the active site (Davies et al., 1969; Segal et al., 1971, 1972).

Further insight into the structural basis of chymotrypsin (alpha and gamma) cleavage specificity has been

---

[2]See Appendix 1 for the abbreviations used to designate different amino acids.

possible by studying the three-dimensional structure of complexes formed in the active site region. However, a fundamental limitation inherent in the X-ray crystallographic technique requires the use of inhibitors or other substrate-like molecules in the study of substrate binding. This arises from the lengthy periods of time (on the order of days) required to collect adequately informative diffraction data, far beyond the time frame upon which enzymatic reactions occur. Nevertheless, at least two strategies (both of which have been applied to chymotrypsin) can permit the visualization of detailed structural information relevant to substrate binding which is unobtainable by other means. One such method, applied to alpha-chymotrypsin, used the virtual substrate N-formyl tryptophan (Steitz et al., 1969). The three-dimensional structure of the N-formyl tryptophan complex in the active site of alpha-chymotrypsin showed that the indolyl side chain lies in a large hydrophobic pocket. This pocket had earlier been shown to bind the toluene ring of the p-toluene sulphonyl inhibited form of alpha-chymotrypsin. These experiments suggest this hydrophobic pocket is responsible for the cleavage point discrimination shown by alpha-chymotrypsin on true substrates.

* Chloromethyl ketone peptide inhibitors bound covalently to gamma-chymotrypsin have also provided insight into substrate binding (Segal et al., 1971, 1972). Crystallographic analysis of such inhibitors have clearly

identified the reactive active site histidine to be His-57. In addition, these inhibitors have revealed binding subsites further removed from the primary specificity pocket, since they have been constructed from a number of amino acids spanning a considerable area on the enzyme surface. These studies indicate that at least four amino acids N-terminal to the scissile bond would lie on the enzyme surface before further amino acids of longer substrates would extend into surrounding solvent.

Also synthesized in the bovine pancreas as an inactive precursor, and subsequently activated in the small intestine, is the serine protease trypsin (Keil, 1971). Trypsin has a molecular weight of approximately 23,300 (223 amino acid residues) and cleaves peptide bonds on the carbonyl side of lysine and arginine residues. This enzyme is also responsible for the activation of chymotrypsinogen A by cleaving this zymogen precursor between Lys-15 and Ile-16. The subsequent formation of a salt bridge between the newly formed N-terminal at Ile-16 and the side chain of Asp-194 leads to the active form of chymotrypsin (Hess, 1971).

Despite the completely different cleavage specificities exhibited by alpha-chymotrypsin and trypsin, the high degree of primary sequence homology between these two enzymes, suggested they had very similar tertiary structures. This has been subsequently shown to be the case as a result of the structural determination of trypsin (Stroud et al.,

1974; Fehlhammer and Bode, 1975; Bode and Schwager, 1975). These studies have also shown that a similar 'catalytic triad' or arrangement of a serine, a histidine and an aspartate residue is present in the active site of trypsin.

Further comparison of the structures of the primary specificity pockets of these two enzymes, shows them to be similarily constructed except for the replacement of Ser-189 at the bottom of this pocket in alpha-chymotrypsin for an aspartate residue in trypsin. Thus, the primary specificity pocket of trypsin is ideally suited for long substrate side chains with a positively charged end; such as lysine or arginine, which could interact with Asp-189.

Also isolated from bovine pancreas is a potent protein inhibitor of trypsin. This inhibitor functions to inactivate prematurely activated trypsin molecules. Such premature activation of trypsin, if unchecked, could lead to further activation of other zymogen precursors, such as chymotrypsinogen A, before they are transported to their proper sites of activity. Both the three-dimensional structure determination of bovine pancreatic trypsin inhibitor (EPTI) and the inhibitor complex formed with trypsin have been completed (Deisenhofer and Steigemann, 1975; Ruhlmann et al.,1973; Huber et al., 1974). These studies show that a surface loop of BPTI lies in the trypsin active site, much as short peptide inhibitors of chymotrypsin lie on the surface of that enzyme. A lysine side chain of BPTI was found bound in the primary

specificity pocket of trypsin. Of even more interest, these studies have shown that the strong inhibitor activity of BPTI apparently arises not only from its similarity to true substrates but also from its ability to form a stable intermediate near the active site residues of trypsin. This intermediate is much like transistory species postulated to cccur during true substrate catalysis.

Further crystallographic structural studies of the zymogen precursors of alpha-chymotrypsin and trypsin have lead to the elucidation of the conformational changes involved in the activation of these enzymes (Fehlhammer et al., 1977; Freer et al., 1970; Wright, 1973; Birktoft et al., 1976; Kossiakoff et al., 1977). Both chymotrypsinogen A and trypsinogen differ from their active enzymatic forms in that neither have been selectively cleaved, leading to the generation of a free N-terminus at Ile-16. Thus, in these zymogen structures the salt bridge between Ile-16 and Asp-194 is absent. Upon zymogen activation, there is a dramatic repositioning of the side chain of Asp-194 and the new N-terminal residue Ile-16 to form a salt bridge. This appears to promote the solidification of polypeptide chains in the primary specificity pocket region. These portions of polypeptide chain are flexible in the zymogen structures. It is interesting to note that the conformations of active site residues in both zymogens is similar to those found in the activated enzymes. Thus zymogen activation is dependent on the reorientation of the side chains of Asp-194 and Ile-16

as well as the solidification of the substrate binding region rather than upon a major realignment of catalytic residues.

Less extensive structural studies have been carried out on a third serine protease, also synthesized as an inactive precursor in the mammalian pancreas. Porcine pancreatic elastase (approximate molecular weight 25,900; 240 amino acid residues) is specific for the cleavage of peptide bonds on the carbonyl side of the amino acids alanine and valine (Hartley and Shotton, 1971). Like trypsin, elastase has a significantly different specificity from that of alpha-chymotrypsin although it retains a marked degree of primary sequence homology with alpha-chymotrypsin . The tertiary structure of elastase has also been elucidated by crystallographic methods (Shotton and Watson, 1970; Sawyer et al., 1978). These analyses show elastase to have a similar overall tertiary structure to that found for trypsin and alpha-chymotrypsin. The specificity of elastase for small amino acid side chains can be explained by the presence of two side chains not present in alpha-chymotrypsin. The side chains of Thr-226 and Val-216, block the primary binding cleft reducing its effective size to a shallow surface pocket capable of accommodating only small amino acid side chains.

The high sequence homology evident in comparisons of the amino acid sequences of alpha-chymotrypsin, trypsin and elastase is even more pronounced about the active site

residues of these enzymes. In particular, there is complete conservation of the active site sequence Gly-Asp-Ser-Gly-Gly about the reactive serine residue. This has led to the designation of these enzymes as being of the Asp-Ser-Gly family of serine proteases. The term 'family' here is taken to mean a group of functionally similiar enzymes that are sequentially homologous and similar in overall three-dimensional conformation.

Another distinct family of serine proteases, thus far derived exclusively from bacterial sources, has also been studied extensively. These bacterial enzymes are known collectively as the subtilisins and have the active site sequence Thr-Ser-Met about their reactive serine residue. Nevertheless, subtilisins exhibit a similar reactivity with both serine and histidine directed reagents such as diisopropyl fluorophosphate and chloromethyl ketone peptides as do the pancreatic serine proteases (Kraut, 1971). This demonstrated that subtilisins also utilize the reactive side chain of a histidine and of a serine residue in the catalytic process. However, in spite of an apparently similar catalytic mechanism and function of these enzymes, there is no primary sequence homology between the subtilisins and the pancreatic family of serine proteases (Kraut, 1971).

The structural relationship between the subtilisin and pancreatic serine protease families has been elucidated by X-ray crystallographic studies of subtilisin BPN' and

subtilisin Novo. These two enzymes, both of whose structures have been independently solved, were once thought to be related isozymes but have since been shown to be identical (approximate molecular weight 27,500; 275 amino acid residues)[3] (Kraut, 1977). As expected from the initial lack of primary sequence homology, structural studies have shown that the subtilisins are folded in a very different manner than the pancreatic serine proteases (Wright et al.,1969; Drenth et al., 1972). However, two pancreatic-like features are conserved in the subtilisin structure. The first of these is a similar juxtaposition of a serine, a histidine and an aspartate residue in the active site, a common feature in all serine proteases for which structures have been resolved. Secondly, peptide chloromethyl ketone studies, similar to those conducted with gamma-chymotrypsin show that the substrate binding region of subtilisin bears a marked similarity to that found for the pancreatic serine proteases (Robertus et al., 1972a). Thus, subtilisins or Thr-Ser-Met serine proteases, while protraying completely different overall primary and tertiary structures, retain the essential elements of the serine protease proteolytic mechanism. Therefore, the pancreatic and subtilisin families of enzymes appear to represent a good example of the convergent evolution of a common catalytic mechanism from unrelated ancestral genes.

---

[3]In the present work, this enzyme is referred to simply as subtilisin.

Investigation into the catalytic mechanism of serine proteases, particularly that of alpha-chymotrypsin, has had a long history. By using a large variety of probes, the basic sequence of events occurring during the cleavage of peptide bonds has been established. Not only does the catalytic process appear to be identical for the family of serine proteases homologous with alpha-chymotrypsin, but also for the subtilisin family. This is not surprising since the three-dimensional conformation of catalytic residues in all serine proteases has been found to be nearly identical. Unfortunately these studies are too numerous to discuss in depth in this work, but excellent reviews have been presented by Bender and Killheffer (1973) and by Kraut (1977). The latter review incorporates more recent structural studies.

Shown in Figure 1 are the catalytic residues of elastase, whose conformation is representative of that found for serine proteases (Sawyer et al., 1978). The aspartate residue of the so called catalytic triad of residues is buried and isolated from solvent contact. Besides the hydrogen bond interaction formed to the catalytic histidine residue, the aspartate residue forms additional hydrogen bonds with other enzyme groups. The catalytic histidine residue is on the surface of the enzyme near the side chain of the reactive serine. Earlier structural studies had suggested that a hydrogen bond was formed between the histidyl and seryl side chains. It was also postulated that

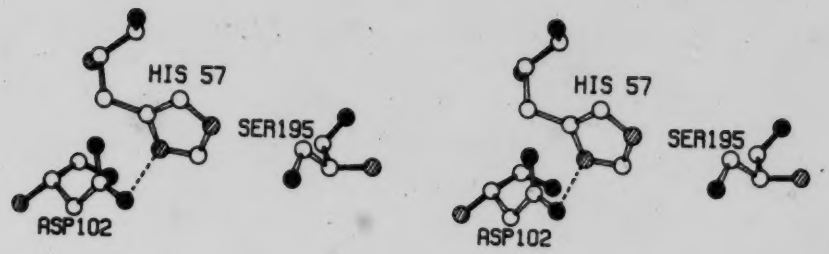


Fig. 1. Stereo-drawing of the three-dimensional configuration of Asp-102, His-57 and Ser-195 in the active site of porcine pancreatic elastase. The polypeptide main chain bonding is shown with solid black bonds and oxygen atoms are distinguished by solid black circles. Striped circles indicate nitrogen atoms. The interaction between Asp-102 and His-57 is shown by a dashed line.

the Asp-His couple induced the formation of a nucleophilic alkoxide ion on the side chain of the active serine. However, subsequent high resolution structural studies of a number of serine proteases, indicate that there is little or no interaction between the histidyl or seryl side chains in the resting state of the active site (Matthews et al., 1977). It is now believed the catalytic serine hydroxyl group derives its nucleophilicity by being ideally poised to interact with the carbonyl carbon of a bound susceptible peptide bond (Kraut,1977). The Asp-His couple is seen as a mechanism for the transfer of a proton from the attacking serine hydroxyl to the amide nitrogen of the substrate leaving group.

A schematic representation of the proposed catalytic mechanism of peptide bond cleavage by serine proteases is

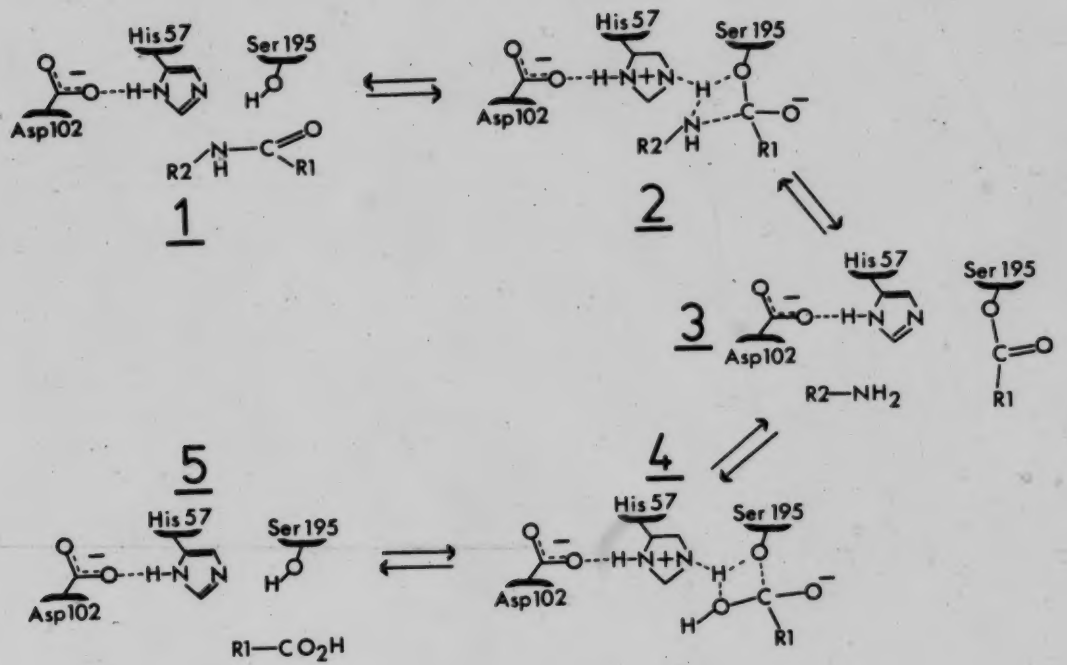


Fig. 2. A schematic representation of the proposed catalytic mechanism of serine proteases. This drawing is based on similar illustrations in Polgar and Bender (1969), and Kraut (1977).

shown in Figure 2 (Polgar and Bender, 1969; Kraut, 1977). Cnly the catalytic triad of enzyme side chains are shown in Figure 2 and in this scheme R1 and R2 represent further atoms of the substrate not specifically depicted, but that are connected across the susceptible peptide bond. As can be seen, the proposed reaction sequence is symmetrical, consisting of a number of steps leading to acylation of the enzyme (intermediate 3) followed by a similar deacylation process (Kraut, 1977).

The first intermediate of catalysis shown in Figure 2, is the non-covalent Michaelis complex or the initial approach of a substrate into the active site. This is

followed by the formation of a tetrahedral complex in which a covalent bond is formed between the reactive serine gamma oxygen atom and the carbonyl carbon atom of the susceptible peptide bond. At this point, the hydroxyl hydrogen atom of the reactive serine is in the process of being transferred to the leaving group amide via the catalytic histidine residue. With the completion of this transfer the tetrahedral intermediate breaks down to the acyl enzyme (intermediate 3) liberating the free leaving group.

Subsequent deacylation of the enzyme is simply the reverse of the acylation process. Complexation of a water molecule with the acyl enzyme leads to the formation of a second tetrahedral intermediate (intermediate 4 in Figure 2). The hydroxyl portion of water forms a covalent bond to the carbonyl carbon atom of the cleaved peptide and the remaining hydrogen atom is in the process of being transferred via the catalytic histidine residue to the reactive serine side chain. With the transfer of a hydrogen atom to the catalytic serine residue the final product of catalysis is released (intermediate 5) with a terminal planar carboxyl group and subsequently leaves the enzyme surface. The catalytic triad of the enzyme is now ready to catalyze the next susceptible substrate peptide bond bound.*

---

*Serine proteases also catalyze the hydrolysis of ester bonds. Although peptide and ester bond cleavages are believed to proceed via the same overall mechanism, the rate limiting step for each is different. For peptide bonds the conversion of intermediate 2 to intermediate 3 is rate limiting while for ester bonds the rate limiting step is the conversion of intermediate 3 to intermediate 4 (Fersht, 1977).

Although the pathway of peptide cleavage has been established, the actual mechanism of serine proteases, that is the peculiar property of these enzymes responsible for the phenomenonal enhancement of catalytic activity, is still in some question. More recently, transition state theory, based on earlier proposals of Pauling (1948), has been evoked to explain the catalytic activity of enzymes such as serine proteases (Wolfenden, 1972; Lienhard, 1973). The basic philosophy of this theory rests on the premise that the catalytic activity of an enzyme is due to the preferential binding of a substrate molecule in a configuration characteristic of its activated transition state complex. Thus the enzyme surface is seen as a physical constraint that selectively binds the substrate so that the susceptible chemical bond to be altered approaches its transition state geometry thereby lowering the activation energy required to catalyze the reaction. In this view an enzyme catalyzes a particular reaction because it is a template for binding the transition state complex of that reaction.

Several structural features of the active site surfaces of serine proteases are held in common, and thus can be identified as physical constraints in the binding of substrates in a suitable manner. These features include:

1. An extended polypeptide binding site on the acyl side of the susceptible peptide bond.
2. A number of well developed binding sites for the side chains of the polypeptide substrate.

3. A site for binding the carbonyl oxygen atom of the susceptible peptide bond when the carbonyl group is in a planar or tetrahedral configuration (known as the oxyanion hole; Robertus et al., 1972b).
4. A reactive serine residue ideally positioned to form a covalent bond with the carbonyl carbon atom of the susceptible peptide bond.

An example of the ability of the surface of a serine protease to distort a susceptible bond can be seen in the bovine pancreatic trypsin inhibitor-trypsin complex (Ruhlmann et al., 1973; Huber et al., 1974). The peptide bonds of uncomplexed BPTI have normal planar conformations. However, upon complexation of this inhibitor with trypsin, in a manner much as expected for a true substrate, the peptide bond bound near the catalytic triad is tetrahedrally deformed. It has been further shown that tetrahedralization is due to enzyme-inhibitor surface contacts rather than solely from interactions formed with catalytic residues.

A logical consequence of transition state theory is that inhibitors, which have features like those of the transition state of a true substrate, will be bound tightly in the active site. This has been demonstrated for serine proteases with the use of inhibitors having the ability to form tetrahedral adducts such as boronic acids, sulphonyl or phosphoryl fluorides and aldehydes. Structural studies show that such inhibitors form stable covalent tetrahedral

complexes in the active sites of serine proteases, not unlike similar transient species postulated to occur during normal substrate catalysis.

B. The Extracellular Serine Proteases Of *Streptomyces griseus* (Strain K1)

The non-motile, gram-positive microorganism known as *Streptomyces griseus*, is isolated from soil and river muds (Buchanan and Gibbons, 1974). This bacterial organism is characterized by a mycelial vegetative structure (grey in color) which is analogous to that which occurs in fungi. Reproduction takes place by the formation of special cells known as conidiospores which detach from the mycelium of mature colonies and are capable of giving rise to the germination of a new mycelial colony (Stanier *et al.*, 1970).

Industrial cultivation of *Streptomyces griseus* (strain K1) was initiated after it was discovered that the antibiotic streptomycin could be isolated from cultures of this organism. Interest in the proteolytic enzymes of the K1 strain of *Streptomyces griseus* began with the observation that a remarkable amount of proteolytically active material was excreted by this microorganism during the production of streptomycin (Nomoto and Narahashi, 1959a). This proteolytic component was normally destroyed under the severe conditions employed during conventional procedures used in streptomycin purification. Nomoto and Narahashi (1959a) devised a method of recovering both the protease and antibiotic components,

via a series of successive ultracentrifugation and precipitation steps. Further purification of the proteolytic component, which these investigators termed pronase, was achieved by chromatographic methods.

Initial investigation of pronase indicated it had an unusually broad specificity (Nomoto et al., 1960a,b) despite its characterization as a homogeneous enzyme (Nomoto and Narahashi, 1959b). Indeed, pronase liberated virtually all amino acids in protein degradation experiments. Such an extremely broad specificity led to considerable speculation about whether pronase was in fact a single protease or in reality a complex mixture of many proteolytic enzymes. The first successful fractionation of pronase into a number of distinct proteolytic components, was accomplished by Hiramatsu and Ouchi (1963). These authors were able to demonstrate the presence of at least four proteolytically active components, separated using the starch zone electrophoresis method.

The development of a variety of different chromatographic techniques has yielded an even more complete breakdown of the components of pronase. For example, Narahashi et al. (1968) provide evidence for the presence of eleven different enzymatically active components by using three different chromatographic columns; Lofqvist and Sjoberg (1971) detected thirteen active components using polyacrylamide gel electrophoresis; and Jurasek et al. (1971) found six proteolytic components using CM-sephadex.

The pronase fractionation system of Jurasek et al. (1971) is of particular importance to the present work, since the extracellular Steptomyces griseus enzymes for which tertiary structures have been resolved were isolated in this manner. This fractionation procedure uses ion-exchange chromatography on a CM-sephadex column and a linear gradient of a volatile buffer (pyridine-acetic acid) at low pH, rather than the previously more common method using CM-cellulose (Trop and Birk, 1970; Narahashi et al., 1968; Wahlby, 1969). In this way, an improved resolution of enzymatically active components in sufficient quantities for their characterization, could be obtained. Nevertheless, the autodigestion of some enzymatic components of pronase is expected using this procedure, since it is carried out in the absence of calcium ions (Lofqvist and Klevhag, 1974; Nomoto et al., 1960b).

It has now been established that pronase contains at least four serine proteases (Trop and Birk, 1970; Gertler and Trop, 1971; Awad et al., 1972), two or more neutral proteases (Narahashi et al., 1968; Lofqvist and Klevhag, 1974), at least two amino peptidases (Narahashi et al., 1968; Vosbeck et al, 1973) and a carboxypeptidase (Narahashi et al., 1968; Lofqvist and Klevhag, 1974). The best characterized of these enzymes are the serine proteases. Three of these were recognized as belonging to the chymotrypsin-trypsin family of serine proteases, based on the detection of the polypeptide sequence Asp-Ser-Gly in

their active sites (Wahlby and Engstrom, 1968). The fourth serine protease appears to belong to the subtilisin family of bacterial serine proteases (Awad et al., 1972). However, this designation is based solely upon peptide cleavage patterns and the presence of the subtilisin active site sequence Thr-Ser-Met has not been conclusively demonstrated.

The discovery of bacterial serine proteases in the extracellular filtrate of Streptomyces griseus (strain K1) and their characterization as Asp-Ser-Gly types generated considerable interest. Up to this point, it had been assumed that all bacterial serine proteases would be of the subtilisin type. It should be noted that just prior to this another bacterial serine protease of the Asp-Ser-Gly type had been isolated from Myxobacter 495. Thus, such proteases are not simply restricted to the genus Streptomyces. The isolation of these bacterial enzymes raised the possibility that they may represent evolutionary precursors of the mammalian pancreatic serine proteases (Olson et al., 1970). This has resulted in numerous studies and comparisons of the microbial and mammalian Asp-Ser-Gly proteases. Of particular interest is the evolution of the catalytic mechanism of serine proteases and the development of structural features related to substrate specificity and in the zymogen activation phenomenon.

Since a number of research groups have participated in the isolation and elucidation of the properties of the three Asp-Ser-Gly serine proteases of Streptomyces griseus, a

variety of different names have been assigned to each enzyme. In this work, the naming convention of Johnson and Smillie (1971) and of Jurasek et al. (1974) is used. Thus the three pronase serine proteases are herein referred to as Streptomyces griseus Protease A, Streptomyces griseus Protease B, and Steptomyces griseus Trypsin or by their abbreviations: SGPA; SGPB and SGT, respectively. SGPA has been labelled by other investigators as: alkaline serine proteinase a (Narahashi and Yoda, 1977); PNPA-hydrolase I (Wahlby, 1969); lysine-free chymoelastase (Siegal and Awad, 1973); Streptomyces griseus enzyme II (Gertler and Trop, 1971) and Streptomyces griseus protease 3 (Bauer and Lofqvist, 1973). SGPB has been designated as: alkaline serine proteinase c (Narahashi and Yoda, 1977); PNPA-hydrolase II (Wahlby, 1969); Streptomyces griseus enzyme III (Gertler and Trop, 1971); guanidine-stable chymoelastase (Siegal and Awad, 1973) and Streptomyces griseus protease 1 (Bauer, 1978). SGT has been described as: alkaline serine proteinase b (Narahashi and Yoda, 1977); BAEE-hydrolase (Wahlby and Engstrom, 1968) and pronase trypsin (Trop and Birk, 1970). The structural studies making up the bulk of this dissertation are centered around the enzymes SGPA and SGPB. Thus further discussion is largely restricted to these two enzymes.

The substrate specificity of SGPA has been examined in several laboratories and found to exhibit protease and esterase activity of wide specificity. For example, SGPA

shows common substrate cleavage characteristics with not only bovine alpha-chymotrypsin, but also with porcine elastase, two mammalian enzymes with widely different specificities. Thus, SGPA hydrolyzes such typical alpha-chymotrypsin substrates as PNPA (Wahlby, 1969) and ATEE (Gertler and Trop, 1971; Johnson and Smillie, 1971) as well as the elastase-like substrate Ac-Ala-Ala-Ala-ME (Gertler and Trop, 1971; Bauer and Lofqvist, 1973). A study of the cleavage pattern of the oxidized A and B insulin chains by SGPA, also concluded that this enzyme showed cleavage activity over a wide range of amino acids (Johnson and Smillie, 1971). Understandably, these initial studies led to considerable confusion in the assignment of substrate specificity to SGPA.

It was not until a more systematic analysis of cleavage specificity had been carried out (Bauer et al., 1976a,b; Bauer, 1978) that it was realized binding subsites further removed from the active site of SGPA also play an important part in the catalytic process. Thus, SGPA is a serine protease for which the definition of the amino acids flanking the scissile bond does not provide a complete picture of those features of the substrate leading to rapid hydrolysis. For example, Bauer et al. (1976a,b) have shown that the basic primary specificity of SGPA is for cleavage on the carbonyl side of the side chains of phenylalanine, tyrosine and leucine. However, cleavage can also occur at the peptide bonds of much smaller amino acids if the

substrate peptide tested is long enough to form suitable interactions remote from the active site. Substrate length dependence studies indicate that important substrate-enzyme interactions are distributed over 6-7 binding subsites in the active site region (Bauer, 1976). Those surface binding subsites that contribute to increased hydrolysis rates by SGPA include subsites S4 through S3'.[5]

The substrate specificity of SGPB has also been extensively studied and found to be very similar to that of SGPA. The systematic comparison of the hydrolysis rates of N-carbobenzoxy amino acid p-nitro phenyl esters and of N-benzoyl amino acid ethyl esters, shows preferential cleavage occurs when the amino acid bound in the primary specificity site is a phenylalanine, tyrosine or leucine residue (Narahashi, 1972). Further analysis of the cleavage patterns of the B chain of oxidized insulin, angiotensin II and oxytocin also indicate hydrolytic activity is directed towards peptide bonds involving the carbonyl groups of large hydrophobic residues. SGPB like SGPA, shows a much broader specificity towards other amino acid residues than alpha-chymotrypsin (Narahashi and Yoda, 1973). The importance of interactions formed in binding subsites of

---

[5]The subsite labelling scheme of Schechter and Berger (1967) is used throughout this manuscript. According to this scheme the enzyme binding site is partitioned into a series of subsites. The residue on the C-terminal end of the bond being cleaved is termed P1 and subsequent residues in the N-terminal direction of the bound peptide are called P2, P3 and so on. In a similar fashion peptide residues bound on the N-terminal side of the scissile bond are referred to as P1', P2', etc. The portion of the enzyme binding PX is then called the SX binding subsite of the enzyme.

SGPB further removed from the active site has also been shown in a study using a series of chloromethyl ketone peptide inhibitors (Gertler, 1974). For example, the chloromethyl ketone Tos-Phe-CK does not inhibit SGPB, whereas the inhibitor Boc-Gly-Leu-Phe-CK shows high activity.

Recent studies (Bauer et al., 1976a,b; Bauer, 1978) have compared the substrate specificity of SGPA, SGPB and alpha-chymotrypsin using a series of synthetic peptide amides as substrates. These studies also demonstrate that both SGPA and SGPB exhibit very similar broad peptide cleavage specificities. However, both enzymes showed only poor hydroylsis rates when the side chain of tryptophan was bound in the primary specificity site. In contrast, alpha-chymotrypsin cleaves at appreciable rates only when the amino acid bound in the primary specificity site is phenylalanine, tyrosine or tryptophan. SGPA and SGPB also show a strong substrate length dependence which is only weakly manifested in alpha-chymotrypsin. These authors conclude that both SGPA and SGPB are unable to orient the scissile bond of a bound substrate properly, when enzyme-substrate interactions are limited to the S1 binding subsite. This is illustrated by the inability of these enzymes to hydrolyze N-acetyl amino acid amides. Nevertheless, such limited primary specificity pocket interactions formed on the surface of alpha-chymotrypsin are sufficient to result in appreciable catalytic activity.

These results suggest that the primary specificity pocket of both microbial enzymes is less developed in terms of providing a surface upon which to orient the scissile bond than is their pancreatic counterpart. Also, the inability of the microbial enzymes to cleave at tryptophan residues indicates their primary specificity site is probably smaller than that of alpha-chymotrypsin. It is apparent from substrate length dependence studies, that the microbial enzymes have compensated for a lack of specific binding in the primary binding site by employing interactions in subsites further removed from the active site in correctly positioning the scissile bond. The development of very specific binding in the primary specificity site of alpha-chymotrypsin means that this enzyme does not require such orientational information and not surprisingly shows less substrate length dependence than the microbial enzymes.

The complete primary sequences of SGPA and SGPB have been determined. SGPA was found to consist of a single polypeptide chain of 181 amino acids of molecular weight 18,012 (Johnson and Smillie, 1974; L.B. Smillie and P. Johnson, personal communication). Two disulfide bridges, linking residues 42 to 58 and 191 to 220, are present. This enzyme is unusual in that it contains no lysine residues; a fairly common amino acid in the sequences of other serine proteases. The alignment of the primary amino acid sequence of SGPA, with those of alpha-chymotrypsin and elastase,

indicates there is good sequence homology in the region of the three catalytic residues Asp-102, His-57 and Ser-195. This was used as a strong indication that the catalytic mechanism and conformation of catalytic residues in SGPA and the pancreatic enzymes would also be very similar. Nevertheless, SGPA has only 18% overall primary sequence identity with either of elastase or alpha-chymotrypsin (Delbaere et al., 1975). This is in contrast to the 39% primary sequence identity found between elastase and alpha-chymotrypsin. Thus, the much smaller overall size of SGPA, coupled with the lack of significant sequence homology, indicated the possibility of significant structural differences being present between this microbial enzyme and the pancreatic family of Asp-Ser-Gly serine proteases.

The subsequent elucidation of the primary sequence of SGPB clearly demonstrated that this enzyme is a close homologue of SGPA, with which it has 59% primary sequence identity. The SGPB molecule consists of 184 amino acids in a single polypeptide chain having a molecular weight of 18,635 (Jurasek et al., 1974; L.B. Smillie and L. Jurasek, personal communication). Two disulfide bridges are present and are similarily placed as in SGPA. SGPB, like SGPA, has significant primary sequence homology with elastase and alpha-chymotrypsin only about catalytic residues in the active site. Overall SGPB has approximately 20% and 17% sequence identity with elastase and alpha-chymotrypsin

respectively (Delbaere et al., 1975). As was the case with SGPA, this low overall primary sequence homology between SGPB and the mammalian pancreatic serine proteases, brought into question the structural relationship between the microbial and pancreatic types of Asp-Ser-Gly serine proteases.

Although few studies of the catalytic mechanism of SGPA and SGPB have been done, the available information indicates that they function in a manner similar to the mammalian enzymes. For example, both enzymes are inhibited by diisopropyl fluorophosphate (Wahlby and Engstrom, 1968). The primary amino acid sequence about the reactive serine residue involved, Gly-Asp-Ser-Gly-Gly, is identical with that found in alpha-chymotrypsin, trypsin and elastase. Although neither microbial enzyme is inhibited at an appreciable rate by the chymotrypsin inhibitor TPCK (Johnson and Smillie, 1971; Gertler and Trop, 1971), it has been demonstrated that longer peptide chloromethyl ketones are effective inhibitors of SGPB (Gertler, 1974). In addition, the single histidine residue of SGPB alkylated by these inhibitors has a neighbouring primary sequence similar to that found about the catalytic histidine residue of alpha-chymotrypsin.

Further evidence of mechanistic similarity between the microbial and pancreatic Asp-Ser-Gly proteases is provided in pH dependence studies. SGPA catalyzed hydrolysis of PNPA and glutaryl phenylalanine p-nitroanilide is dependent on

the ionization of a group with an apparent pKa of 6.6 (Bauer and Pettersson, 1974). Similarily, SGPB hydrolysis of Ac-Pro-Leu p-nitroanilide is dependent on the ionization of a group of apparent pKa 6.7 (Bauer, 1977). Also, the inactivation rate of SGPB by Ac-Leu-Phe-CK shows an apparent pKa of 6.6 (Gertler, 1974). These results agree well with the pKa of approximately seven observed in pH-dependence studies of the pancreatic serine proteases (Bender and Killheffer, 1973). Also, kinetic studies of SGPA hydrolysis of the non-specific ester substrate PNPA and the specific peptide substrate glutaryl phenylalanine p-nitroanilide are consistent with a three step mechanism of Michaelis complex formation followed by enzyme acylation and deacylation (Bauer et al., 1974) as has been postulated for the mammalian enzymes (Figure 2). Thus, despite a notable lack of sequence homology; inhibitor, pH-dependence and kinetic studies suggest a common catalytic mechanism is utilized by both the microbial enzymes SGPA and SGPB, and by the mammalian pancreatic serine proteases.

C. The Extracellular Serine Protease Of Myxobacter 495

Serious interest in certain soil bacteria began in the early nineteen-sixties. It was discovered that a number of these cultures could exert strong attractive forces on nematodes, even though these latter organisms could not derive food nor propagate on the attractant bacterial cultures (Katznelson and Henderson, 1962, 1964). Some of the

bacteria feeding nematode strains attracted (Caenorhabditis briggsae, Rhabditis oxycerca and Panagrellus sp.) were degraded by the extracellular cell-free fluid extract of the bacterial cultures (Katznelson et al., 1964). The most intensively studied of these bacterial soil cultures was that of Myxobacter 495.[6] This gram-negative microorganism characteristically forms rods and filaments (Christensen and Cook, 1978). Since the formation of fruiting bodies is not observed, it is believed that reproduction by this organism occurs only by cell division. An interesting feature of this bacterial culture is its ability to glide along solid-liquid interfaces and to show flexing movements in liquid medium, even though flagella are not present.

Two major proteolytic enzymes (alpha and beta lytic protease) were isolated from the extracellular filtrate of Myxobacter 495. These enzymes were adsorbed from the cultural filtrate on Amberlite CG50 and subsequently displaced from the resin with citrate buffer containing a gradient of sodium citrate concentration (Whitaker, 1965). The two extracellular enzymes of Myxobacter 495 are capable of the degradation of some nematodes, the complete lysis of various species of Staphylococcus, Bacillus, Sarnia and Arthrobactor, and the partial lysis of bacteria from several other genera (Gillespie and Cook, 1965; Whitaker et al., 1965a). Although the classification of the zinc containing

---

[6] It has recently been proposed that this microorganism be reclassified as Lysobacter enzymogenes (Christensen and Cook, 1978)

beta lytic protease is uncertain, alpha lytic protease was shown to be a serine protease with the active site amino acid sequence Asp-Ser-Gly (Whitaker et al., 1966; Whitaker and Roy, 1967). Thus, alpha lytic protease belongs to the same family of serine proteases which contain all known mammalian serine proteases for which polypeptide sequences have been determined. In this regard, alpha lytic protease is similar to the microbial serine proteases SGPA and SGPB isolated from Streptomyces griseus (strain K1). The isolation of these three enzymes firmly establishes the existence of a class of microbial serine proteases related to the mammalian pancreatic serine proteases and distinctly different from the Thr-Ser-Met family of bacterial serine proteases.

Cleavage specificity studies of alpha lytic protease have shown it to have many properties in common with porcine elastase. Analysis of oxidized insulin A and B chain cleavage patterns (Whitaker et al., 1965b) and a systematic comparison of esterase activities, show that alpha lytic protease preferentially cleaves on the carbonyl side of small neutral L-amino acids (Kaplan and Whitaker, 1969; Kaplan et al., 1970). Esters of alanine were the best substrates although valine esters were moderately good. However, esters of glycine, leucine, isoleucine and D-alanine were very poor substrates. Alpha lytic protease, like elastase, also shows a marked preference for long substrates, indicating the presence of secondary binding

subsites further removed from the scissile bond on the enzyme surface (Whitaker *et al.*, 1965b). This enzyme is also capable of hydrolyzing the oligo-glycyl cross-linkages at the C-termini of the peptide chains in the *Micrococcus* mucopeptide (Tsai *et al.*, 1965). Such hydrolytic activity is believed to be partially responsible for the swelling and resultant lysis of bacterial cells placed in extracellular filtrates of *Myxobacter* 495.

The elucidation of the complete polypeptide sequence of alpha lytic protease revealed 198 amino acids in a single polypeptide chain (Olson *et al.*, 1970). This enzyme contains three intra-chain disulfide bridges and has an overall molecular weight of 19,869. The alignment of the primary amino acid sequence of alpha lytic protease with those of pancreatic elastase and alpha-chymotrypsin, revealed there is only approximately 19% and 18% primary sequence identity, respectively (Delbaere *et al.*, 1975). Significant regions of sequence homology were found only near the catalytic residues of elastase and alpha-chymotrypsin. Thus from sequence alignment studies alone, the structural relationship between alpha lytic protease and the pancreatic serine proteases was difficult to discern.

In this regard, alpha lytic protease is similar to the two microbial serine proteases isolated from *Streptomyces griseus* (strain K1). It is interesting to note that alpha lytic protease has 35% and 36% primary sequence identity with SGPA and SGPB respectively (Delbaere *et al.*, 1975).

This suggests alpha lytic protease, SGPA and SGPB are more closely related structurally than they are with their pancreatic counterparts and strongly supports the possibility that these microbial serine proteases arose from a common ancestral gene.

Like SGPA and SGPB, alpha lytic protease appears to catalyze peptide bond cleavage via a mechanism similar to that of the mammalian pancreatic serine proteases. This is reflected not only in the preservation of common polypeptide sequences about the three amino acid residues believed to be intimately involved in the catalytic event, but also in the susceptibility of this enzyme to the common serine protease inhibitor diisopropyl fluorophosphate (Whitaker and Roy, 1967). Both pancreatic elastase and alpha lytic protease share a common specificity and are found to be resistant to derivatization by chloromethyl ketone peptides (Kaplan and Whitaker, 1969). Kinetic studies of the hydrolysis of N-acetyl valine methyl ester and of p-nitrophenyl trimethyl acetate are consistent with the general catalytic mechanism derived for the pancreatic serine proteases (Figure 2). Also in common with other serine proteases, the catalytic activity of alpha lytic protease is dependent on the ionization of a group with a pKa of approximately 6.7 (Kaplan and Whitaker, 1967).

By virtue of having only one histidine residue in its polypeptide sequence, that being the active site residue His-57, alpha lytic protease has played an important role in

the elucidation of the active site mechanism for serine proteases. For example, the single histidine residue of alpha lytic protease (Smillie and Whitaker, 1967; Kaplan and Whitaker, 1969; Kaplan et al., 1970) led to the demise of earlier proposals for the involvement of two histidine residues in the catalytic event of serine proteases (Walsh et al., 1964; Bender and Kedzy, 1964). The presence of only one histidine residue has also made alpha lytic protease the subject of intensive NMR analyses as a probe of the catalytic mechanism (Robillard and Shulman, 1974a,b; Hunkapiller et al., 1973; Bachovchin and Roberts, 1978). Although these studies in conjunction with related kinetic analyses have led to insight into the common catalytic mechanism of serine proteases, a considerable controversy still surrounds basic details of the catalytic event (Kraut, 1977).

D. Objectives Of This Research Project

The discovery of a class of microbial serine proteases related to the mammalian pancreatic serine proteases generated considerable interest, as prior to this, it had been assumed that all bacterial serine proteases would be of the subtilisin type. Based on the sequences of short fragments from the active site region, it was initially suggested these newly isolated bacterial enzymes were representatives of the evolutionary precursors of the pancreatic enzymes. However, further attempts to delineate

the nature of the relationship between the pancreatic and pancreatic-like microbial enzymes, lead to considerable confusion. Even though, the newly discovered microbial enzymes shared with the pancreatic enzymes short stretches of active site sequences, overall there was very little sequence homology indeed. Furthermore, the microbial proteases exhibited rather unusual cleavage specificities. These differences, coupled with the much smaller size of the bacterial enzymes, made the overall relationship between the pancreatic and the microbial enzymes difficult to discern.

A methodology potentially capable of resolving this dilemma is that of X-ray crystallography. That is, by determining the three-dimensional structures of the microbial enzymes at sufficiently high resolution, comparisons of basic tertiary structure can be made with the pancreatic enzymes. Such structural studies could determine if the microbial enzymes exhibit a similar polypeptide chain folding pattern as the pancreatic enzymes, thereby suggesting a evolutionary relationship between these two groups of enzymes. If this were the case, comparisons could further show how certain structural elements have evolved to their present mammalian form. Of special interest in this regard, is the development of the active site region and the facility for zymogen activation. Comparison studies of this type would also be able to pinpoint structural features essential to the catalytic process, as these would be conserved in enzymes from widely different sources.

Of particular interest are the active sites of the pancreatic-like serine proteases. Structural studies could show the configuration of active site residues and the placement of substrate binding sites. In this regard, the structure of alpha lytic protease is of special interest, as this enzyme has become the focus of attempts to understand the mechanism of catalysis of the serine proteases. Also, such studies could potentially explain the peculiar specificity exhibited towards substrates by SGPA and SGPB.

By utilizing the X-ray crystallographic technique, the present study has attempted to define the structures of the pancreatic-like microbial serine proteases, as well as, lend some insight into puzzling aspects of their behavior. To this end, the three-dimensional structure determinations of SGPA, SGPB and alpha lytic protease have been initiated. The crystallographic experiments undertaken to attain these goals are the topic of the following chapters.

Protein crystallography, perhaps more than other investigative techniques, is a team effort and the contributions of my collegues to this work are gratefully acknowledged. In summary, the steps in the determination of the crystal structure of an enzyme include: crystal growth, characterization, intensity data collection, heavy-atom derivative preparation, Patterson solution, heavy-atom coordinate refinement and phase determination, electron density map interpretation, model building and measuring, and finally interpretation of the resulting structure in

terms of its biological significance. My unique contributions have been made at almost all of these stages for the enzymes SGPA and alpha lytic protease. K. Hayakawa was instrumental in producing crystals and determining the optimal heavy-atom soaking conditions. A. Sielecki assisted in the model building calculations. The chapter on the SGPB structure does not include data concerning the structure solution of this enzyme, since I was not involved in the intial stages of this work (the reader is referred to Delbaere et al., 1975). My contributions were in the model building stages and in the interpretation and comparisons with other structures. The inhibitor studies discussed herein were done in collaboration with C.-A. Bauer (synthetic tetrapeptide aldehyde), L. Delbaere (chymostatin) and A. Gertler (chloromethyl ketone peptides). Finally, I wish to give special recognition to L. Delbaere and M. James for advice and encouragement at all stages of the present work.

II. Amino Acid Sequence Alignment Of The Bacterial and Pancreatic Serine Proteases

The first complete amino acid sequence determination of a pancreatic-like bacterial serine protease was that of alpha lytic protease from Myxobacter 495 (Olson et al., 1970). Subsequent analyses revealed the primary sequences of SGPA (Johnson and Smillie, 1974) and SGPB (Jurasek et al., 1974) from Streptomyces griseus (strain K1). In order to gain insight into the relationship between these microbial enzymes and their pancreatic counterparts, a number of investigators have attempted to align the microbial and pancreatic primary sequences. The first such comparison, between alpha lytic protease and elastase (Olson et al., 1970), chose as anchor points, the three disulfide bridges 42 to 58, 168 to 182 and 191 to 220. Two of these disulfide bridges, 42 to 58 and 191 to 220, were considered equivalent in the two enzymes because of their occurrence in highly homologous sequence regions containing active site residues. However, the third disulfide bridge, 168 to 182, occurred in a region with relatively little sequence homology. Overall, the resulting sequence alignment of alpha lytic protease with elastase required large insertions and deletions.

When the primary structures of the bacterial proteases from Streptomyces griseus became known, neither of these enzymes had a disulfide bridge at position 168 to 182 as did alpha lytic protease. Nevertheless, the alignment originally derived for alpha lytic protease with elastase (subsequently

slightly modified by a model building attempt (McLachlan and Shotton, 1971)) was adhered to (Johnson and Smillie, 1974; Jurasek et al., 1974). As found for alpha lytic protease, the maintenance of anchor points at positions 168 and 182, when aligning the sequences of SGPA and SGPB with those of the pancreatic enzymes, also results in the large deletions and insertions. A further sequence alignment was proposed by Delbaere et al. (1975), based on the preliminary structure determination of SGPB. However this alignment differed only marginally from those presented previously and was also unable to produce a good fit of the microbial and pancreatic sequences.

Due to the poor overall sequence homology between the microbial and pancreatic serine proteases, and the tentative nature of their primary sequence alignments, a great deal of confusion has existed over the assignment of amino acid residue numbers when comparisons of these enzymes are made. In order to rectify this problem and to present a consistent numbering scheme upon which to discuss the tertiary structures of SGPA, SGPB and alpha lytic protease, a primary sequence alignment based on the three-dimensional structural comparison of the microbial and pancreatic serine proteases has been carried out. This analysis used the tertiary structures of SGPA, SGPB and alpha lytic protease; the three structures which form the basis of the present dissertation, as well as those of alpha-chymotrypsin (Birktoft and Blow, 1972) and elastase (Shotton and Watson, 1970). The tertiary

structure of trypsin was not examined in this manner due to an inability to obtain reliable coordinates. Nevertheless, the sequence alignment of trypsin with the microbial enzymes is also presented based on its primary sequence homology to alpha-chymotrypsin and elastase.

In the formation of the present primary sequence alignment, the original numbering scheme of chymotrypsinogen A (Hartley and Kauffman, 1966) for alpha-chymotrypsin has been preserved as have the sequence alignments of elastase and trypsin to alpha-chymotrypsin. Therefore, all adjustments to the primary sequence alignment of the microbial enzymes to their pancreatic counterparts have been made to the microbial primary sequences.

Of the six polypeptide sequences examined, four minor modifications were made to the published sequences: (a) Ser-60 of SGPA was deleted (L.B. Smillie and P. Johnson, personal communication); (b) for SGPA Asn-123 has been redesignated as Asp-123 (A. Sielecki and L.B. Smillie, personal communication); (c) for SGPB, Ala-68 has been reinterpreted as Trp-68 and Val-186N has been added, based on the interpretation of the 2.8 angstrom resolution electron density map of this enzyme (Delbaere et al., 1975); (d) Gln-70 and Gln-80 of bovine trypsin have been reinterpreted as Glu-70 and Glu-80 (Bode and Schwager, 1975).

Structural comparisons of SGPA, SGPB, alpha lytic protease, alpha-chymotrypsin and elastase were carried out in a pairwise fashion by determining the topological

equivalence between each pair of enzymes as established using a program (written by W.S. Bennett) based on the proposals of Rossmann and Argos (1975). That is, determining which atoms of each enzyme occupy the same relative space in the tertiary structures of the pair of enzymes. In this procedure, only the alpha-carbon atoms of the amino acid residues of each enzyme were used.

Initial rotation and translational matrices relating the tertiary structures of two different enzymes were obtained using twelve equivalent alpha-carbon atom positions in both molecules. The alpha-carbon positions chosen were from those few regions of high sequence homology in all the serine protease structures examined. These included the alpha-carbon positions of residues 56 to 58, 101 to 103, 193 to 196 and 214 to 215, the numbering of which has not changed in the new sequence alignment. Upon the computation of initial parameters, the relative orientation and translation of all alpha-carbon positions were then refined in alternate cycles of least-squares minimization of the distances between equivalent atoms followed by the reassignment of the equivalences. A progression rule by which the equivalences must be chosen sequentially was used. This procedure was iterated until the best fit of the two enzymes being compared was achieved. The topological equivalences thus determined, represent the starting point for the sequence alignment presented in Table 1. Overlap stereo-drawings were then prepared for each pair of enzymes.

The regions of topological equivalences which resulted from the comparison program were readily evident in these drawings and they further assisted in the preparation of the sequence alignment. Table 1 also indicates those amino acids conserved in all the primary sequences aligned and regions of topological equivalence.

The first part of the present alignment is not significantly different from those presented earlier (Olson et al., 1970; Johnson and Smillie, 1974; Jurasek et al., 1974). However, from position 112 to 190 in the present table of sequences, significant rearrangement of the bacterial alignment to agree with the tertiary structural comparisons has resulted. Whereas there were a large number of deletions in the bacterial sequences from residues 113 to 168 (present numbering) relative to the pancreatic enzyme sequences, there is now only a single major deletion at positions 144 to 155 and no large insertions. Adjustment of the primary sequence in this area has also led to the reassignment of the anchor point disulfide bridge 168-182 (old numbering) of alpha lytic protease that had been earlier equated to the disulfide bridge 168-182 of the pancreatic enzymes (old and new numbering). The earlier presumption, that the third disulfide bridge of alpha lytic protease had a counterpart in the pancreatic enzymes was responsible for the large number of insertions and deletions in the original sequence alignments between residues 112 and 190 (old and new numbering). It is apparent from the present

TABLE 1

Amino Acid Sequence Alignment of Streptomyces griseus Protease A (SGPA), Streptomyces griseus Protease B (SGPB), Alpha Lytic Protease (a-LP), Alpha-Chymotrypsin (CHYM), Elastase (ELAS) And Bovine Trypsin (BT)

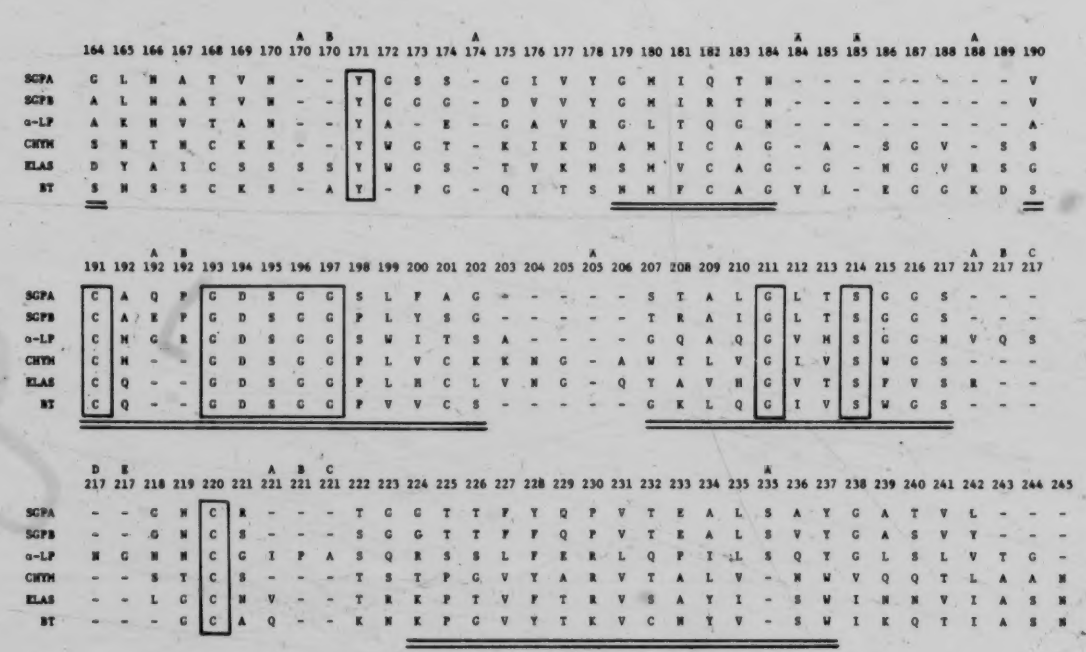
	14A	14B	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36A	36B	36C	37	38	39	40	41	42	43
SGPA	-	-	I	A	G	G	-	-	-	-	-	-	-	-	-	-	E	A	I	T	T	G	-	-	-	-	-	-	-	G	S	R	C	S
SGPB	-	-	I	S	G	G	-	-	-	-	-	-	-	-	-	-	D	A	I	Y	S	S	-	-	-	-	-	-	-	T	G	R	C	S
a-LP	A	N	I	V	G	G	-	-	-	-	-	-	-	-	-	-	I	E	Y	S	I	N	N	-	-	-	-	-	-	A	S	L	C	S
CHYM	-	-	I	V	N	G	E	E	A	V	P	G	S	W	P	W	Q	V	S	L	Q	D	K	-	-	-	T	G	F	H	F	C	G
ELAS	-	-	V	V	G	G	T	E	A	Q	R	N	S	W	P	S	Q	I	S	L	Q	Y	R	S	G	S	S	W	A	H	T	C	G
BT	-	-	I	V	G	G	Y	T	C	G	A	N	T	V	P	Y	Q	V	S	L	N	-	-	-	-	-	S	G	Y	H	F	C	G

	44	45	46	47	48	48A	48B	48C	48D	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65A	65	66	67	68	69	70	71
SGPA	L	G	F	N	V	S	V	N	G	V	A	N	A	L	T	A	G	H	C	T	-	-	N	I	S	A	S	W	-	-	-	-	-
SGPB	L	G	F	N	V	R	S	G	S	T	Y	Y	F	L	T	A	G	H	C	T	D	-	G	A	T	G	T	W	W	-	-	-	-
a-LP	V	G	F	S	V	T	R	G	A	T	K	G	F	V	T	A	G	H	C	G	-	-	T	V	N	A	T	A	R	-	-	-	-
CHYM	G	S	L	I	N	-	-	-	-	E	N	W	V	V	T	A	A	H	C	G	V	T	T	S	D	V	-	V	V	A	G	E	F
ELAS	G	T	L	I	R	-	-	-	-	Q	N	W	V	N	T	A	A	H	C	V	D	R	E	L	T	F	R	V	V	V	G	E	N
BT	G	S	L	I	N	-	-	-	-	S	Q	W	V	V	S	A	A	H	C	Y	K	S	G	I	Q	V	R	L	-	-	G	E	D

	72	73	74	75	76	77	78	79	80	80A	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	99A	99B	100	101
SGPA	-	-	-	-	-	-	-	-	-	-	-	-	-	S	I	G	T	R	T	G	T	-	S	F	-	-	-	-	-	P	-	N	N
SGPB	-	-	-	-	-	-	N	S	A	-	R	T	T	V	L	G	T	T	S	G	S	-	S	F	-	-	-	-	-	P	-	N	N
a-LP	-	-	-	-	-	-	-	-	I	-	G	G	A	V	V	G	T	F	A	A	R	-	V	F	-	-	-	-	-	P	-	G	N
CHYM	D	Q	G	S	S	S	E	K	I	-	Q	K	L	K	I	A	K	V	F	K	N	S	K	Y	N	S	L	T	I	-	-	N	N
ELAS	N	L	N	Q	N	N	G	T	E	-	Q	Y	V	G	V	Q	K	I	V	V	H	P	Y	W	N	T	D	D	V	A	A	G	Y
BT	N	I	N	V	V	E	G	N	E	-	Q	F	I	S	A	S	K	S	I	V	H	P	S	Y	N	S	N	T	L	-	-	N	N

	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	120A	120B	120C	120D	121	122	123	124	125	126	127	128	129	130
SGPA	D	Y	G	I	I	R	N	S	N	P	A	A	A	D	G	R	V	Y	L	Y	N	G	S	Y	Q	D	I	T	T	A	G	N	A
SGPB	D	Y	G	I	V	R	Y	T	N	T	T	I	P	K	D	G	T	V	G	-	-	-	-	G	Q	D	I	T	S	A	A	N	A
a-LP	D	R	A	W	V	S	L	T	S	A	Q	T	L	L	P	R	V	A	N	-	G	S	S	F	V	T	V	R	G	S	T	E	A
CHYM	D	I	T	L	L	K	L	S	T	A	A	S	F	S	Q	T	V	S	A	-	-	-	-	V	C	L	P	S	A	S	D	D	F
ELAS	D	I	A	L	L	R	L	A	Q	S	V	T	L	N	S	Y	V	Q	L	-	-	-	-	G	V	L	P	R	A	G	T	I	L
BT	D	I	N	L	I	K	L	K	S	A	A	S	L	N	S	R	V	A	S	-	-	-	-	I	S	L	P	T	-	S	C	A	S

	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163
SGPA	F	V	G	Q	A	V	Q	R	S	G	S	T	T	-	-	-	-	-	-	-	-	-	-	-	-	G	L	R	S	G	S	V	T
SGPB	T	V	G	N	A	V	T	R	R	G	S	T	T	-	-	-	-	-	-	-	-	-	-	-	-	G	T	H	S	G	S	V	T
a-LP	A	V	G	A	A	V	C	R	S	G	R	T	T	-	-	-	-	-	-	-	-	-	-	-	-	G	Y	Q	C	G	T	I	T
CHYM	A	A	G	T	T	C	V	T	T	G	W	G	L	T	R	Y	T	N	A	N	T	P	D	R	L	Q	Q	A	S	L	P	L	L
ELAS	A	N	N	S	P	C	Y	I	T	G	W	G	L	T	R	-	T	N	G	Q	L	A	Q	T	L	Q	Q	A	Y	L	P	T	V
BT	-	A	G	T	Q	C	L	I	S	G	W	G	N	T	K	S	S	G	T	S	Y	P	D	V	L	K	C	L	K	A	P	I	L



Note: The numbering is that of chymotrypsinogen A (Hartley and Kauffman, 1966), with insertions in the sequences of the enzymes shown being denoted as 15A, 15B, etc. Deletions are denoted by broken lines. Those 17 residues that are chemically identical in all six protein sequences are enclosed by solid lines. The residues that are doubly underlined are those that are topologically equivalent in these serine proteases. The single letter code for amino acids is used in this table.

tertiary structure comparisons, that the third disulfide bridge of alpha lytic protease does not have a homologous counterpart in the pancreatic enzymes or for that matter in the other microbial enzymes. In Table 1 this disulfide bridge is indicated as being between positions 137 and 159.

A second major change in the primary sequence alignment has occurred in the region of residues 164 to 182 (new numbering). In previous alignments this region was thought to have a 17 or 18 residue insertion in the bacterial enzymes' sequences and was termed by Shotton and McLachlan (1971) as the 'extra compensating loop'. This loop was

believed to compensate for the reduced size of neighbouring polypeptide loops as defined by earlier sequence alignment attempts. However, from structural comparisons it is now clear that this large loop is simply a reorganized tertiary structural component of the pancreatic enzymes' methionine loop (residues 164 to 182). The methionine loop is so named because of the presence of a methionine residue at position 180 in this loop in all the pancreatic enzymes (Table 1). As can be seen in Table 1, from position 190 to the C-terminal, essentially the same alignment as in previous publications has been maintained.

Table 2 presents a convenient summary of the primary sequence alignment of Table 1 and the topological equivalence results found for the six proteases compared. The upper triangular portion of the matrix of Table 2 shows that the bacterial enzymes SGPA, SGPB and alpha lytic protease are highly homologous in sequence amongst each other, but have only a small portion of identical residues when compared with the pancreatic enzymes (maximum is 21% for SGPA and alpha-chymotrypsin). Another distinction between the bacterial proteases and the pancreatic enzymes with regard to topological equivalence can be observed in the lower triangular portion of Table 2. Greater than 80% of the alpha-carbon atom positions of the three bacterial enzymes are topologically equivalent. However, in comparisons of the tertiary structures of these enzymes with those of alpha-chymotrypsin and elastase, this equivalence

TABLE 2

Amino Acid Sequence Identity And Topological Equivalence Matrix For The Proteins of Table 1

	SGPA (181)	SGPB (185)	a-LP (198)	CHYM (230)	ELAS (240)	BT (223)
SGPA	-	111(61)	64(35)	39(21)	33(18)	38(21)
SGPB	154(85) 1.46	-	66(36)	33(18)	35(19)	31(17)
a-LP	148(82) 1.46	154(83) 1.76	-	35(18)	36(18)	39(20)
CHYM	116(64) 1.96	117(63) 2.07	114(58) 2.05	-	94(41)	101(45)
ELAS	106(59) 1.76	117(63) 2.15	108(55) 2.02	208(86) 1.02	-	87(39)
BT	-	-	-	-	-	-

Note: The total number of residues in each protein is given in parentheses below its name in the heading. The upper triangular portion of this matrix contains the number of chemically identical residues (the percentage of the number of residues in the smaller is in parentheses) for each pair of proteins as aligned in Table 1. The lower triangular portion of the matrix has the results of the topological equivalence comparisons. For each pair of proteins compared, the number of topologically equivalent residues (the precentage is in parentheses) and the root mean square deviation in angstroms is given. For enzyme abbreviations, see Table 1.

is reduced to around 60%.

The present realignment of the bacterial serine protease primary sequences, while in some regions being dramatically altered from earlier alignments, is similar to previous alignments in the small number of amino acid residues found conserved between the pancreatic-like bacterial proteases and those isolated from the mammalian pancreas. Clearly, the initial confusion of earlier

investigators in aligning these sequences is understandable. These results serve to point out the tenuous nature of alignments of primary sequences from distantly related sources having little sequence homology and few solid anchor points about which to assist the alignment process.

It is remarkable that despite the very low sequence identity between the microbial and pancreatic serine proteases that these enzymes share approximately 60% topological equivalence. Unlike earlier sequence alignments, these results strongly suggest that the microbial and pancreatic serine proteases are structurally related. The significance of structural features held in common by these enzymes, and those that are not, are discussed more fully in following chapters. Note that the alignment and numbering scheme of Table 1 is used exclusively in following discussions.

III. Molecular Structure Of *Streptomyces griseus* Protease A At 2.8 Angstrom Resolution

A. Isolation And Crystallization

SGPA was isolated from the commercial preparation, pronase (Grade B, Calbiochem, lot 801930), as described by Jurasek *et al.* (1971). Crystals of purified SGPA were obtained by the technique of equilibrium dialysis (Zeppezauer *et al.*, 1968) from solutions of 1% SGPA (w/v) and 1.3M NaH(2)PO(4) at pH 4.1. Well formed crystalline prisms (many with tetragonal pyramidal ends) of suitable size for diffractometer data collection were obtained within a month. The dimensions of these crystals were typically 0.4mm x 0.4mm x 0.6mm and in general they were elongated along the unique axis (Figure 3). Figure 4 shows an h0l precession photograph of a native enzyme crystal. The crystallographic symmetry and unit cell parameters found for native SGPA crystals are summarized in Table 3.

Only one SGPA molecule per asymmetric unit is expected from the observed unit cell parameters and crystallographic symmetry. One can calculate a Vm value of 2.30 cubic angstroms per dalton (Matthews, 1968) for SGPA crystals, assuming a molecular weight of 18,097 for SGPA (Johnson and Smillie, 1974), a unit cell volume of 1.67×10^5 cubic angstroms and that there are four molecules per unit cell. This value for the volume per unit molecular weight is close to the overall mean value of 2.37 cubic angstroms per dalton

Fig. 3. Photomicrograph of crystalline Streptomyces griseus protease A (40X) from 1.3M NaH(2)PO(4) at pH 4.1. The long axis of the crystals is the c axis and the a and b axes are normal to the prominent side faces.

(median value 2.61 cubic angstroms per dalton) found for a variety of protein crystals. Based on these results it is expected that 53% of the volume of SGPA crystals is occupied by protein.

B. Heavy-atom Derivative Preparation

Heavy-atom derivatives were prepared by soaking native SGPA crystals in the respective solutions made up with 1.5M NaH(2)PO(4) at pH 4.1. Preliminary screening of heavy-atom derivatives (a total of 23 were tried) was done photographically on a Nonius precession camera with Cu K-alpha radiation from an Elliott rotating anode generator operated at 40kV and 40mA. Native and derivative precession

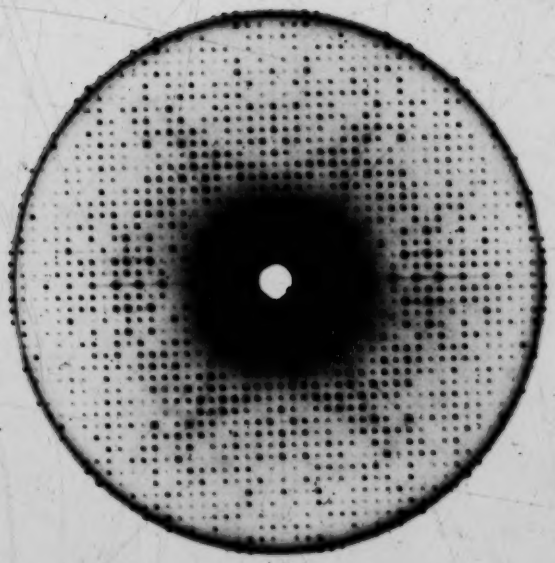


Fig. 4. Precession photograph of a native SGPA crystal showing the h01 diffraction plane to a limit of 2.4 angstrom resolution. This photograph was taken using Ni-filtered Cu K-alpha radiation, 40kV, 40mA and 6 hours exposure time.

photographs were then compared visually for diffraction intensity changes. Heavy-atom compounds that were only slightly soluble in the phosphate buffer were also tried and in one case (mercuric chloranilate) resulted in a suitable derivative. In all, only four suitable heavy-atom derivatives were found (Table 4) and subsequently used to solve the native structure of SGPA.

TABLE 3

Crystal Data For SGPA

Unit cell dimensions	
a	55.14(1) angstroms
b	55.14(1)
c	54.81(2)
V	1.67×10^5 (angstroms)^3
Unit cell content	4
Systematic absences	001:l=2n$\pm$1
Space group	P4(2)
Growth conditions	1.3M NaH(2)PO(4), pH 4.1

C. Data Collection

All crystals were mounted in thin walled glass capillaries approximately 1.0mm in diameter. For the native and each heavy-atom derivative crystal, diffraction intensity data were collected on a Picker FACS-1 diffractometer. The diffractometer computing and controlling system of Lenhert (1975) was used throughout data collection. A Picker X-ray generator with a 0.75mm x 15mm focal-spot copper target tube, was operated at 40kV and 26mA. The incident radiation was Ni-filtered and the diffracted beam was passed through a helium-filled tube extending from near the crystal to the counter. The data were collected with the crystal to counter distance set at 65.0cm. The ambient temperature during data collection was maintained at 15°C, since it was observed that more rapid

decay of crystal reflectivity resulted at higher temperatures.

Unit cell dimensions for the native and heavy-atom derivative crystals were determined from the centered two-theta positions of six reflections in the range 19° < two-theta < 25° and the positions of their minus two-theta Friedel pair mates. Changes in the unit cell dimensions of derivative crystals were minimal, indicating a high degree of isomorphism (Table 4). The data were collected by omega scans of approximately 0.45° in width, using a scan speed of 2°/minute. Four second background counts were measured 0.8° on either side of the center of each reflection along the two-theta direction. The net time spent per reflection was approximately 35s and a complete data set, consisting of about 9100 reflections, required five days to collect.

Crystals of SGPA were sufficiently resistant to radiation damage to allow for the collection of a unique quadrant of diffraction data to 2.8 angstrom resolution. The conditions of heavy-atom soaking and the total number of reflections measured for the native and each derivative crystal are shown in Table 5.

D. Background Correction And Data Reduction

Background measurements during data collection consisted of only two 4s counts on either side of each reflection, in order to keep the total crystal exposure time to a minimum. To compensate for the expected statistical

TABLE 4

Cell Dimension Changes in Derivative Crystals[1]

Data Set	a (sigma a)	%change	c (sigma c)	%change
Native	55.14(1)	-	54.81(2)	-
Mersalyl	55.13(1)	-0.02	54.81(2)	0.00
Mercuric chloranilate	55.10(2)	-0.07	54.86(2)	+0.09
ReCl(3)	55.20(2)	+0.11	54.70(2)	-0.20
ReCl(3) + Mersalyl	55.22(1)	+0.14	54.74(2)	-0.13

[1] a and b were constrained to the same value due to tetragonal crystal symmetry. All unit cell dimensions are in angstrom units. Sigma values represent the precision of a single determination from one crystal.

fluctuation in such short measurements, the intensity data were corrected for background radiation in the following manner. The background counts were fitted by a non-linear least-squares method, with a multi-dimensional function to provide calculated best-fit background counts, which were more reliable than the individual measurements. The sum of the two individual background measurements were examined as a function of the following variables: I (total reflection intensity), two-theta, phi and chi. The background sum was found to be directly proportional to the peak intensity and to a linear combination of two-theta and phi. No chi dependence was observed, contrary to that found by Krieger et al., (1974). Other workers have also found that the background radiation is independent of chi (Hill and

TABLE 5

Statistics of Data Collection to 2.8 Angstrom Resolution For SGPA

Data Set	Native	Mersalyl	Mercury Chloranilate	ReCl(3)	ReCl(3)+ Mersalyl
Soaking time/concentration	-	24h/1mM	3d/sat.	3d/2.5mM	24h/2.5mM, 1mM
Total reflections measured	9165	9159	9189	9150	9157
No. of independent reflections	4117	4117	4119	4124	4127
Maximum absorption correction (%)	14.9	11.7	14.8	9.1	7.3
Maximum crystal decay (%)	10.2	8.2	2.6	5.2	7.0
$R(sym)=(\sum_i \lvert\bar{I}_i - I_i\rvert / \sum_i I_i) \times 100$ (%)	1.7	1.6	2.2	1.6	2.7
No. of reflections merged	4113	577	592	860	860
%Reflections with {I>3(sigma(I))}	96.2	95.0	95.6	95.2	91.2

Banaszak, 1973).

The function (B) used to evaluate the sum of the two backgrounds for any given reflection hkl, with two-theta > 8°, was:

$$B=\{Q1+Q2(I)\}\{1.0+Q3(|TT|)+Q4(TT)^2+Q6\sin^2(phi-Q5)\},$$

where I is the measured peak intensity at the two-theta and phi values of the particular reflection (TT is the two-theta value of that reflection). A weighted least-squares fit of the function was obtained by adjustment of the parameters Q1-Q6. The weights used were the reciprocal of the sum of the individual backgrounds. A modified version of the program BMDX85 (Sampson, 1970) was used for the least-squares computations.

The net intensity for a particular reflection was then:

$$I(NET)=I-(tB)/8$$

and

$$\text{sigma}^2\{I(NET)\}=I+(t/8)^2(\text{sigma}^2(B)),$$

where t is the time spent in scanning the reflection intensity and sigma(B) is the standard deviation of the evaluated background function. Table 6 shows the improvement for three 2.8 angstrom data sets in averaging equivalent reflections after using this function to evaluate the background sum.

The remainder of the intensity data, the shell 4° < two-theta < 8°, did not fit the previously described

TABLE 6

Application of The Background Function

Data set	No. of reflections averaged	R(sym) before	R(sym) after
1	484	2.96	2.25
2	575	2.42	1.84
3	4327	4.64	3.42

background function, probably as a result of the presence of intense background streaking at such low two-theta values. Consequently, this shell of data was processed in the following manner for each individual reflection:

$$I(NET) = I - \{t(B1+B2)/8\}$$

and

$$\text{sigma}^2\{I(NET)\} = I + (0.02)^2 I^2 + (t/8)^2 \{B1+B2+(0.02)^2(B1^2+B2^2)\},$$

where B1 and B2 are the individual measured backgrounds of the reflection and 0.02 is the instrument instability constant.

Once reflection backgrounds had been adjusted, an absorption correction was applied to account for the effects of crystal shape, adhering mother liquor, and the glass capilliary. Absorption was corrected for using the method of North et al. (1968) based on a single averaged absorption curve derived from two 001 reflections (l=6,14). The maximum absorption correction factors were in the range 1.149 to

1.073 for the crystals used in this study (Table 5). Linear decay corrections were determined from the decrease in intensity of three monitor reflections: 3,1,12; 0,12,8; and 12,1,2 measured after every 100 data reflections.

At this point, all symmetry equivalent reflections in the native data set, which consisted of a full set of hkl and -h,-k,-l reflections, were averaged. For heavy-atom derivative data sets, all equivalent reflections other than Friedel mates, contributed mainly from hk0 and kh0 symmetry related reflections, were averaged. The relatively low values of R(sym) shown in Table 5 for each data set, indicate that the data collected from crystals of SGPA and its derivatives are of high quality. Lorentz and polarization corrections were also made and the square roots of the intensities taken in order to derive structure factor amplitudes.

E. Data Scaling

A variant of Wilson's (1942) statistical method was used to determine the absolute scale and overall isotropic thermal parameter of the native data (Thiessen and Levy, 1973). The apparent average isotropic thermal parameter for SGPA is 12.4 (angstroms)^2. This value compares favorably with those found for proteins in a similar molecular weight range: 17 (angstroms)^2 for sea lamprey hemoglobin (Hendrickson *et al.*, 1973), 18 (angstroms)^2 for ribonuclease S (Wyckoff *et al.*, 1970) and 27 (angstroms)^2 for

alpha-chymotrypsin (Tulinsky et al., 1973).

The absolute scale and isotropic thermal parameters for each derivative data set were also determined. The derivative data were then scaled to the native data using the relation:

$$F(PH) = \{scale(D)/scale(N)\} \times F(PH\ unscaled)$$

where scale(N) is the absolute scale value determined for the native data. Scale(D) is the absolute scale determined for the heavy-atom derivative data, and F(PH) is the scaled derivative structure factor amplitude. The absolute scale, overall isotropic B and the ratio (scale(D)/scale(N)) for the native and derivative data sets are shown in Table 7.

Included in the absolute scale calculation for the native protein crystal were all those atoms determined from the amino acid sequence analysis of SGPA (Johnson and Smillie, 1974). On this basis, the molecular formula of SGPA was assumed to be C(774)H(1215)N(229)O(263)S(5). The absolute scale for a derivative data set was calculated assuming one additional fully occupied heavy-atom site per protein molecule. No attempt was made to account for groups that may be attached to the heavy-atoms. Also left out of these calculations were solvent molecules, and in the case of derivative crystals, partially ordered heavy-atoms in solvent regions. However, these molecules contribute mostly to low angle reflections and in order to diminish their effect, only data from 10.0 - 2.8 angstrom resolution were used in the calculation of the absolute scale and the

TABLE 7

Data Set Scaling And Heavy-Atom Differences

Data Set	Native	Mersalyl	Mercury Chloranilate	ReCl(3)	ReCl(3)+ Mersalyl
Wilson plot absolute scale	10.78	10.75	11.29	10.30	18.21
Overall isotropic B (angstroms)^2	12.4	12.6	13.1	11.8	11.5
Abs. scale (D) / Abs. scale (N)	-	0.997	1.047	0.956	1.690
Linear scale ratio	-	0.935	0.987	0.887	1.550
Heavy-atom difference factor R(D)	-	0.073	0.075	0.157	0.209

overall isotropic B value.

The calculated linear scale ratio:

$$\frac{\Sigma F(P)}{\Sigma\{F(PH)^{+} + F(PH)^{-}\}/2}$$

found for each heavy-atom derivative was also determined, using all the data collected. In this ratio $F(PH)^{+}$ and $F(PH)^{-}$ are a Friedel pair of heavy-atom structure factor amplitudes and F(P) is the corresponding native structure factor amplitude. Ratio values determined in this manner are comparable to those determined from the absolute scales of the native and derivative data using only data from 10.0 to 2.8 angstroms.

Following data scaling only those reflections with I > 3sigma(I) were used in subsequent computations. Once a heavy-atom derivative data set was scaled to the native data, the heavy-atom difference factor R(D) was calculated using the equation:

$$R(D) = \frac{\Sigma|\{F(PH)^{+} + F(PH)^{-}\}/2 - F(P)|}{\Sigma F(P)} \quad \text{(sum over all hkl)}$$

Table 7 also shows the linear scale ratio found for each heavy-atom derivative as well as the calculated R(D) value.

F. Heavy-Atom Solution, Phase Calculation And Least-squares Refinement

Three-dimensional difference Patterson maps (Figure 5) were used to derive the x, y coordinates of the heavy-atom sites for the three isomorphous derivatives, mersalyl,

mercuric chloranilate and rhenium trichloride. These maps used:

$$\{(F(PH)^{+} + F(PH)^{-})/2 - F(P)\}^{2}$$

as coefficients and all the observed data to 2.8 angstrom resolution. Fortunately, the first difference Patterson map examined was that of the mersalyl derivative, which was readily interpretable in terms of only one major site. The z coordinate for the single mersalyl site was fixed at a value of 0.0 (P4(2) is a polar space group) and the z coordinates for the other heavy-atom derivative sites were referenced to this site from cross-phased Fourier maps (Dickerson et al.,1967). Such cross-phased Fourier maps were calculated with coefficients:

$$m\{F(H) - F(P)\} \exp(i\ \mathrm{alpha}(P)),$$

where

$$F(H) = \{F(PH)^{+} + F(PH)^{-}\}/2.$$

The figure of merit, m, and the 'best' set (Blow and Crick, 1959) of native phases, alpha(P), were taken from the phase determination carried out with the mersalyl derivative. Coordinates for the three heavy-atom sites of the rhenium trichloride - mersalyl double derivative were determined solely from the results of a cross-phased Fourier map computed from the native protein phases which had been determined from the three previous derivatives.

Three-dimensional anomalous Patterson maps were also calculated for the mersalyl, mercuric chloranilate and rhenium trichloride derivatives using as coefficients

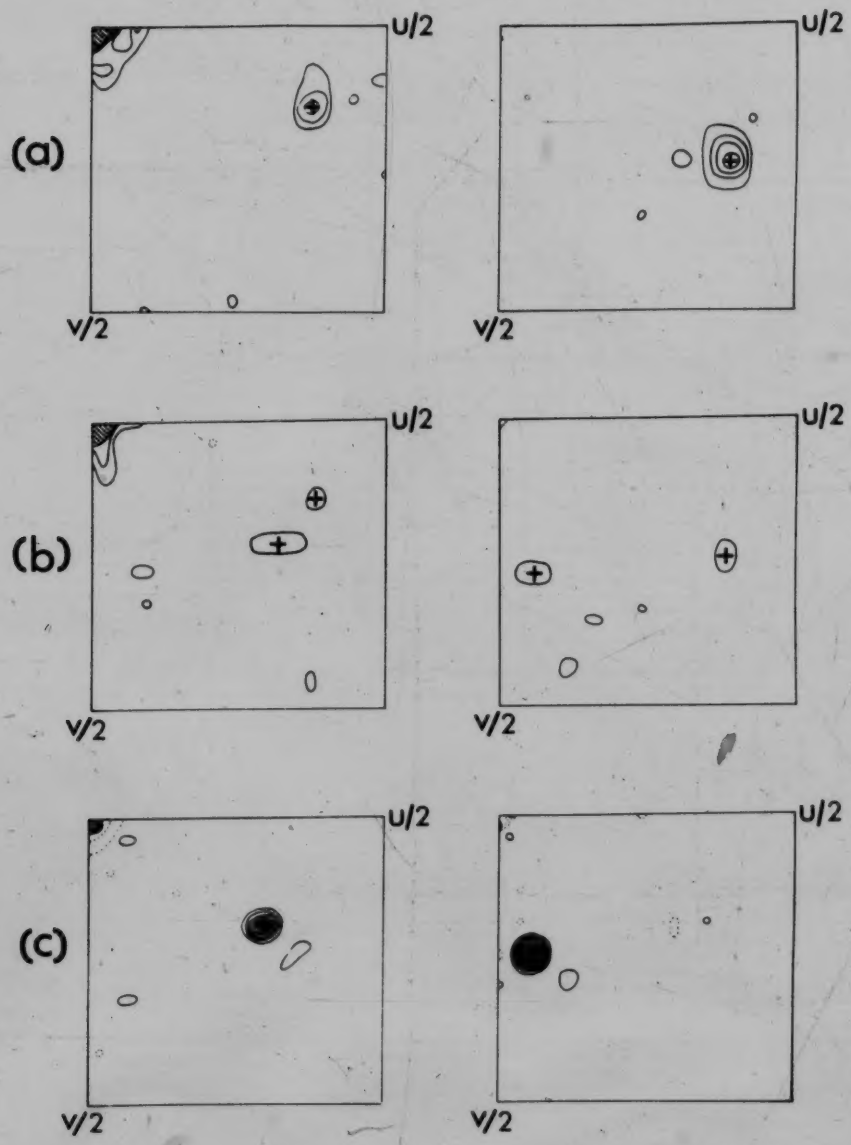


Fig. 5. Heavy-atom difference Patterson maps for the three major derivatives used in this study: (a) mersalyl, (b) mercuric chloranilate and (c) rhenium trichloride. The two Harker sections w = 0 and w = 1/2 are shown for each derivative. Crosses mark the positions of refined heavy-atom sites that were initially determined from these Patterson maps. Only the mercuric chloranilate derivative has more than one heavy-atom binding site.

$\{F(PH)^+ - F(PH)^-\}^2$. Since all the data used to calculate such a map arises from a single crystal, it is not influenced by physical differences between crystals or imprecise crystal to crystal scaling as are the heavy-atom difference Patterson maps. Nevertheless, the anomalous signal is weaker and anomalous Patterson maps have smaller peak to background ratios. A comparison of the anomalous and heavy-atom difference Patterson maps of each derivative indicated both were identical with respect to peak positions. Thus these anomalous Patterson maps confirmed the heavy-atom coordinates determined from the difference Patterson maps.

Lack of closure error refinement (Dickerson *et al.*, 1961, 1968) was used to determine the final phases of the reflections from the native protein crystal. The function minimized in this procedure was:

$$E(H) = \Sigma w\{F(PH) - |\underline{F(P)} + \underline{f(H)}|\}^2 ,$$

where $E(H)$ is the lack of closure error and $\underline{f(H)}$ is the calculated heavy atom structure factor. The weights, w, are the overall value of $1/E(H)^2$ for the particular two-theta range in which the reflection was found. The two-theta range of the reflections (4° - 32°) was divided into eight groups for this purpose. A computer program written by Adams *et al* (1969) was used. This program produces a new set of native phase angles after each cycle of refinement of all

heavy-atom derivative parameters. These new phase angles, alpha(P), and heavy-atom parameters, were then used to determine the vector sum $|\underline{F(P)} + \underline{f(H)}|$ and the lack of closure errors for the next cycle of refinement. The scale factor and the heavy-atom positional parameters were refined in each cycle, but the isotropic thermal parameter and site occupancy were not varied in the same cycle due to the high correlation observed between them. The phase determination process was also modified to include the contribution of anomalous scattering which resulted in reduced lack of closure errors.

Prior to the completion of heavy-atom refinement, a double difference Fourier map was calculated for each derivative using the lack of closure errors that had been found, to determine if minor secondary heavy-atom sites had been overlooked (Blake et al., 1963). Coefficients for this calculation were:

$$\{F(PH) - |\underline{F(P)} + \underline{f(H)}|\} \exp(i\ \text{alpha(PH-calc)}).$$

Alpha(PH-calc) is the calculated phase for F(PH) from the previous cycle of phase determination. These maps were uniformly free of any significant peaks and no additional heavy-atom sites that had not been elucidated from earlier difference Patterson or cross-phased Fourier maps were detected.

The progress and rate of convergence of refinement was followed by examining the parameter shifts and behavior of quantities sensitive to the refinement process as a function

of resolution or as overall values. The most important of these quantities included: (1) the site occupancy and isotropic B; (2) r.m.s. lack of closure errors E(H); (3) r.m.s. calculated heavy-atom structure factor amplitudes f(H) as related to the closure errors; (4) the average figure of merit and (5) the value of the centric R(c) factor (Cullis et al., 1961) for each derivative as a function of resolution.

G. Phasing Results

The final heavy-atom parameters of the four isomorphous derivatives of SGPA are listed in Table 8. The relatively low occupancies and the small number of binding sites observed for each derivative are probably the reasons for the high degree of isomorphism observed between derivative and native SGPA crystals.

The ratio f(H)/E(H) between the r.m.s. calculated heavy-atom structure factor amplitudes f(H) and the r.m.s. lack of closure errors E(H) for each derivative, proved to be the most important factor in judging the progress of heavy-atom refinement and phase determination. The variation of f(H)/E(H) as a function of {sin(theta)/lambda}^{2} for each derivative is shown in Figure 6. For all derivatives the heavy-atom contribution is greater than the lack of closure errors over all resolution ranges.

The variation in the average figure of merit with {sin(theta)/lambda}^{2} is also shown in Figure 6. A slight

TABLE 8

Refined Heavy-Atom Parameters For SGPA

Derivative	Site	x/a	y/b	z/c	A[1]	B[2]
Mersalyl	1	0.4251	0.1877	0.0000	14.4	10.6
Mercuric chloranilate	1	0.4296	0.1894	0.0011	8.1	8.8
	2	0.3506	0.3926	-0.0963	8.0	22.0
ReCl(3)	1	0.1473	0.0942	0.3778	41.5	12.8
ReCl(3) + Mersalyl	1	0.4236	0.1864	0.0005	18.2	16.0
	2	0.1472	0.0944	0.3781	40.6	11.0
	3	0.3525	0.3892	-0.0926	18.8	16.7

[1]A is the site occupancy on an approximately absolute scale in electrons.
[2]B is the isotropic temperature factor coefficient, in units of (angstroms)^2.

drop-off in the figure of merit curve is expected at higher resolution. Reflections measured at high values of two-theta are generally of smaller magnitude and thus most sensitive to experimental errors and imperfect isomorphism.

A histogram of the figure of merit distribution among the measured native reflections is shown in Figure 7. Of all the reflections for which phases were determined, more than 87% have $m \geq 0.5$. For these 3957 native reflections, the average phase angle difference as computed from the equation $|\alpha(max) - \alpha(best)|$ was 11.6°. In this equation, $\alpha(max)$ is the most probable phase and $\alpha(best)$ is the best phase (Blow and Crick, 1959). The overall average figure of merit for SGPA was 0.82 at the end of heavy-atom refinement.

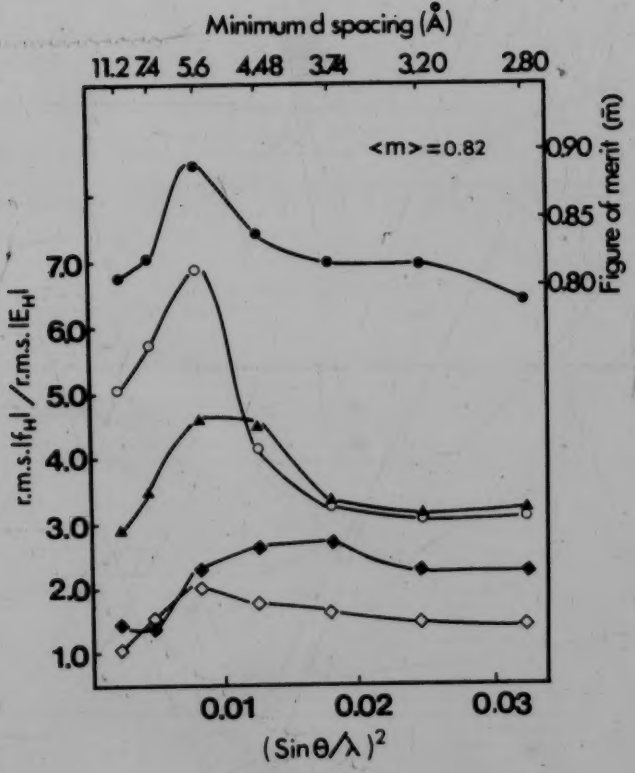


Fig. 6. The variation of the ratio of r.m.s. f(H) to r.m.s. E(H) and the mean figure of merit as functions of {sin(theta)/lambda}². The curves shown represent: (-o-o), FeCl(3); (-◆-◆), mersalyl; (-◇-◇), mercuric chloranilate; (-▲-▲), ReCl(3) and mersalyl. The uppermost curve and the scale to the right show the variation of the mean figure of merit as a function of {sin(theta)/lambda}² (-●-●).

The final phase refinement statistics for SGPA are given in Table 9. The fact that all derivatives showed good phasing power indicates that even derivatives containing only moderately occupied heavy-atom sites, are valuable in the phase determination process.

Another quantity which proved a useful guide to monitor the progress of the heavy-atom refinement was the Cullis

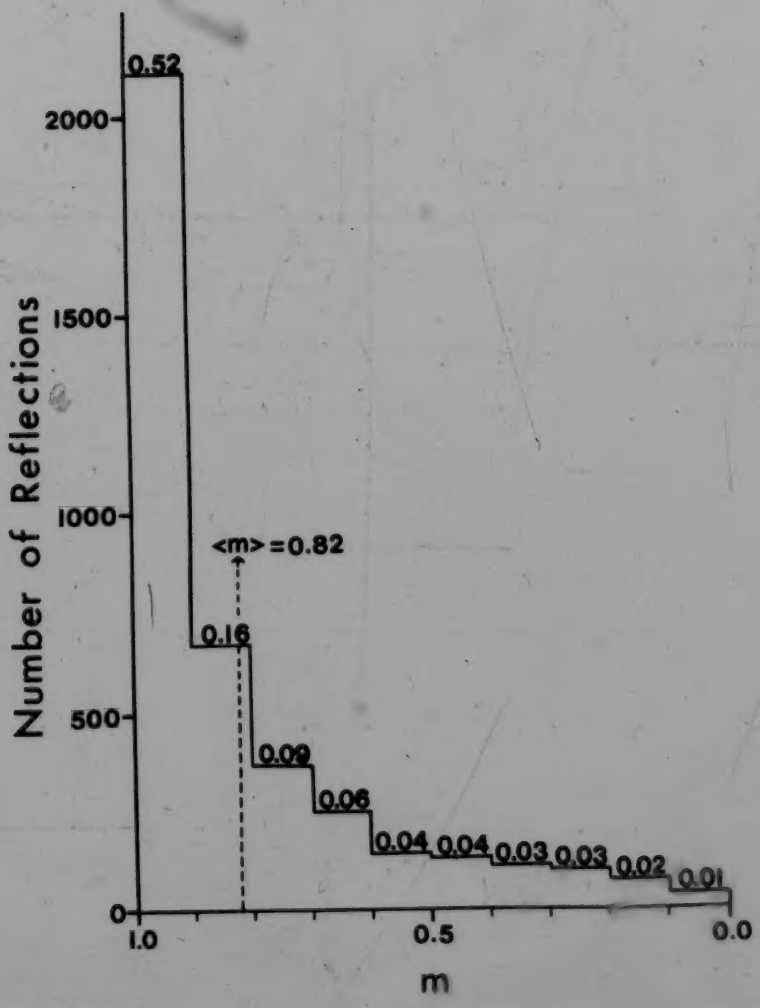


Fig. 7. The distribution of the figures of merit among native enzyme reflections phased using the isomorphous replacement technique. The precentage of the total reflections falling into each range is shown at the top of each column. The broken line indicates the overall mean figure of merit for all reflections.

R(c) factor computed from centric data only (Cullis et al., 1961). The very low values obtained for the rhenium derivatives (Table 9) indicate the high isomorphism of these derivative crystals. The low occupancies of the two binding sites of the mercuric chloranilate derivative and the

TABLE 9

Phase Determination Statistics For SGPA[1]

Derivative	R(D)[2]	R(c)[3]	R(k)[4]	[5]r.m.s. f(H) /r.m.s. E(H)
Mersalyl	0.073	0.432	0.034	2.371
Mercuric chloranilate	0.075	0.623	0.041	1.528
ReCl(3)	0.157	0.339	0.073	3.638
ReCl(3) + Mersalyl	0.209	0.366	0.090	3.424

[1]The overall mean figure of merit was 0.82.
[2]R(D) is the heavy-atom difference R-factor.
[3]R(c) is the Cullis R factor (Cullis et al., 1961).
[4]R(k) is the Kraut R-factor (Kraut et al., 1962).
[5]Over all the reflections phased.

resultant small heavy-atom signal measured, was probably the reason a high R(c) value was observed for this derivative. Nevertheless, the mercuric chloranilate derivative was very valuable in the initial stages of phase determination and significantly improved earlier native protein electron density maps.

The Kraut R factor, shown for each derivative in Table 9, had values varying from 0.034 to 0.090. It was not a particularly sensitive indicator of the progress of heavy-atom refinement (Tulinsky et al., 1973). The overall ratio of r.m.s. f(H)/r.m.s. E(H) for each derivative is also given in Table 9, showing that good phasing power results from all four derivatives used.

H. Heavy-Atom Binding Sites

Only three unique heavy-atom binding sites were observed among the four isomorphous derivatives used to solve the SGPA structure. Heavy-atom to enzyme interactions in these three sites have been studied in detail, using protein atomic coordinates derived from the interpretation of the native enzyme map. The protein groups responsible for heavy-atom binding at each site, along with the interaction distances are given in Table 10.

Soaking native crystals in mersalyl resulted in a single heavy-atom binding site near His-57. There were no negative peaks observed in the double difference maps of this derivative at or near His-57, indicating the position of the imidazole ring had remained unperturbed on heavy-atom binding.

Mercuric chloranilate bound to two sites on the enzyme surface, one of these was the same site as that found for mersalyl. The second site was found in a region later determined to be the specificity pocket of the enzyme. The relatively high thermal parameter and low occupancy of this second site, indicates its poor heavy-atom binding characteristics. The same site had a somewhat higher occupancy in the mersalyl-rhenium trichloride double derivative.

The present study is apparently the first documented example of the use of rhenium trichloride as a heavy-atom derivative in the solution of a protein structure. Rhenium

TABLE 10

Sites of Heavy-Atom Binding

Heavy-atom site	Nearby protein atoms	Distance[1]
Major mersalyl and mercuric chloranilate	Cys-42 SG	4.0
	His-57 NE2	2.5
	His-57 ND1	3.6
	His-57 O	2.7
	Cys-58 SG	3.4
	Cys-58 N	4.1
	Ser-195 OG	4.3
Mercuric chloranilate and double derivative mersalyl	Ala-192 N	3.8
	Ala-192 O	3.1
	Gly-216 O	3.9
	Ser-217 O	3.6
	Gly-218 O	3.4
	Asn-219 N	3.7
Rhenium trichloride	Ser-48A OG	3.0
	Gly-48D O	3.5
	Val-49 O	4.9
	Arg-117 NEH2 (-x,-y,z)	3.7
	Arg-117 NE (-x,-y,z)	4.9
	Tyr-121 N (-x,-y,z)	4.5
	Tyr-121 O (-x,-y,z)	4.7

[1]All distances are in angstrom units.

trichloride binds in a pocket between two enzyme molecules. The chlorine atoms originally bound to the rhenium atom were not detected in either the earlier cross-phased Fourier maps or in the final double difference map produced for this derivative. Indeed, there is insufficient space for these chlorine atoms in the binding pocket, and it appears that the rhenium atom alone is bound to the protein atoms. Normally there are six or eight ligands in the coordination sphere of rhenium (Cotton and Wilkinson, 1972). However, the closest protein interactions to the bound rhenium (Table 10)

do not appear to form a regular coordination sphere. Curiously, the three closest protein interactions appear to form an almost planar trigonal grouping. The other protein-rhenium contacts may have little effect upon rhenium binding as they are greater than 4.5 angstroms in length. Figure 8 illustrates the protein-rhenium contacts found in this derivative crystal, viewed down the c crystallographic axis.

I. Electron Density Maps And Model Building

A Fourier summation, computed with the 'best' phase (Blow and Crick, 1959) and figure of merit for each of the 3957 unique native reflections phased, resulted in a native electron density map. The electron density was sampled at a/76, b/76 and c/54 intervals and then contoured and plotted on a scale of 6mm per angstrom. Plotted sections were photocopied onto plastic sheets and placed on plexiglass sections held 6mm apart. Sections of electron density were then stacked along the crystallographic c axis.

Each enzyme molecule in this map was clearly separated from its neighbours and large solvent regions between molecules extended up to 15 angstroms in width. Naturally, a few close contacts between molecules are present and probably represent the major interactions responsible for maintaining crystal integrity. All heavy-atom sites were located on the enzyme surface.

The standard error in the native electron density map

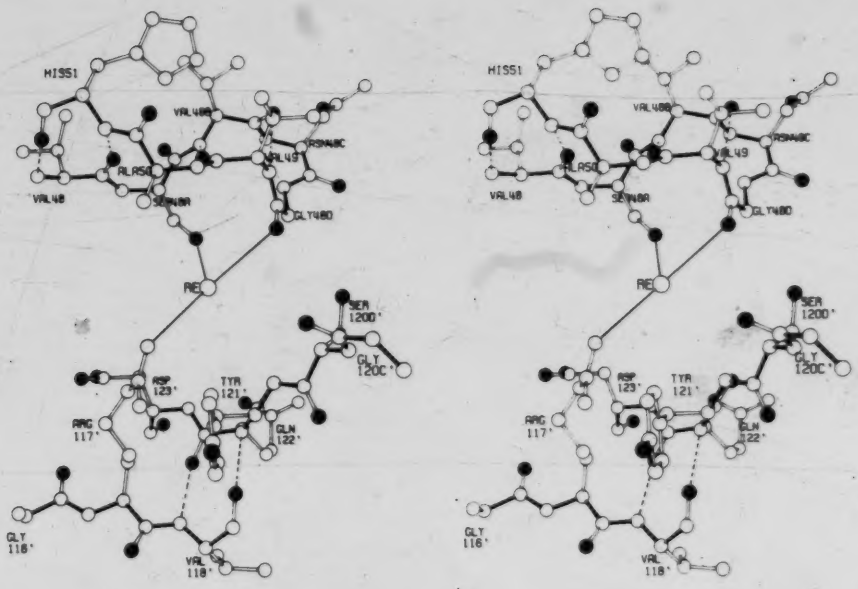


Fig. 8. Stereo-drawing of the rhenium binding site coordination sphere as observed down the c crystallographic axis. The rhenium atom is found in a pocket between two enzyme molecules, one side of which is open to the solvent. Nearby polypeptide chains from each molecule forming the binding pocket are drawn. In addition, all oxygen atoms and main chain bonds have been blacked in. The three most probable protein-rhenium interactions are drawn as thin lines.

was estimated to be 0.172e/(angstroms)^3 (Cruickshank, 1949; Dickerson et al., 1961). This error represents mainly phasing errors, since the standard error due to the measurement of structure amplitudes is only 0.004e/(angstroms)^3. Peak electron densities along the entire length of the polypeptide chain are greater than 3 sigma of the estimated error in the present native electron density map.

The first contour of the native electron density map was drawn at 0.45e/(angstroms)^3 (including 0.23e/(angstroms)^3 contributed by the F(000)/V term). Further contour lines were placed at 0.11e/(angstroms)^3 intervals. The average peak electron density in the native map along the polypeptide chain was approximately 0.87e/(angstroms)^3, and varied from 0.66e/(angstroms)^3 in some surface loops to 1.09 e/(angstroms)^3 in the central region of the enzyme. The largest peaks of electron density occurred at the positions of the two disulfide bridges in the molecule (1.42 and 1.96e/(angstroms)^3).

The chemical sequence of SGPA proved a valuable aid in following the course of the polypeptide chain through the native protein electron density map. Two minor modifications to the original published sequence (Johnson and Smillie, 1974) were taken into consideration in the interpretation of the native electron density. These are that Ser-60 is no longer included and Asn-123 has been redesignated as Asp-123 (A. Sielecki, P. Johnson and L.B. Smillie, personal communication). Also see Table 1 for the revised residue numbering scheme for SGPA. The position of His-57 was discerned from the position of the major mersalyl heavy-atom site. Simple peak height analysis, to determine the most pronounced features of the map was sufficient to locate both the disulfide bridges present in the molecule, as well as the position of the only methionine (180). The position of the only tryptophan in the sequence was determined visually

by inspection of the electron density map for the largest side chain present. With the aid of these marker residues and the chemical sequence, it was possible to follow the course of the entire polypeptide chain, and to determine an approximate alpha-carbon atom position for all 181 residues of the molecule.

The detailed atomic interpretation of the 2.8 angstrom resolution native map was made by plotting a larger scale map (2cm/angstrom) suitable for use in a optical comparator (Richards, 1968). This map was calculated in sections perpendicular to the b crystallographic axis with dimensions: x (-0.56, 0.25); y (-0.26, 0.76); z (0.0, 0.75) and grid intervals of a/76, b/74, c/76. The first contour was drawn to represent an electron density of 0.34e/(angstroms)^3 and subsequent contours were drawn at intervals of 0.11e/(angstroms)^3. This represents a lower contour level than the previous map to ensure that hydrogen bonding patterns would be observed.

Analysis of this map was made by fitting Watson-Kendrew skeletal units, connected to depict the chemical sequence and manipulated into the electron density distribution, to achieve an optimal fit. The approximate alpha-carbon atom positions derived from the earlier smaller map were used as guide coordinates for the construction of the model. Hydrogen bonds, while optimized where possible, were not introduced as fixed constraints in the construction of the model. A valuable aid in following the course of the

polypeptide chain proved to be the carbonyl oxygen peaks, which were especially useful in determining the orientation of individual peptide bonds between amino acids. Many solvent peaks were observed on the exterior surface of the enzyme; however, no attempt was made to interpret this solvent structure.

Coordinates for all non-hydrogen atoms in the SGPA molecule were measured from the model using the plumb-line method. These coordinates were used as guide points in a model building procedure (Diamond, 1966) so as to obtain the best fit of a stereo-chemically correct structure, with standard bond lengths and inter-bond angles, to the measured coordinates. Once the model was sufficiently close to the guide coordinates, some model strain was released by also permitting variation of the inter-bond angle tau{C(alpha)}, the folds of prolines and chi5 in arginines (Diamond, 1974). Variation of omega was not allowed, so that all peptides remained planar. The final r.m.s. deviation of the idealized structure from the original measured coordinates was 0.25 angstroms for the 1265 non-hydrogen atoms in the molecule.

J. Molecular Conformation Of SGPA

Overall SGPA is a globular enzyme with approximate dimensions 44 x 40 x 36 angstroms. The polypeptide chain of SGPA is folded in a manner such that there are two structurally similar hydrophobic domains. The junction of these domains forms a shallow surface depression containing

the active site region. Each domain consists of four anti-parallel beta loops (6-strands of polypeptide chain) hydrogen bonded to produce a barrel-like structure. The four beta loops forming the amino-terminal domain are: the N-terminal (residues 16-41), the histidine (residues 42-58), the uranyl (residues 65A-86), and the aspartate (residues 87-108) loops. The corresponding loops of the carboxyl-terminal domain are: the autolysis (residues 131-163), the methionine (residues 164-182), the serine (residues 195-213), and the specificity (residues 214-228) loops. Similar domains have been described for alpha-chymotrypsin as being distorted cylinders or beta barrels (Birktoft and Blow, 1972). The folded beta sheet and resultant beta barrels in SGPA are structurally similar to those found in alpha-chymotrypsin.

The polypeptide backbone conformation of SGPA is represented in the phi, psi plot of Figure 9. Most of the plotted points in the so-called unallowed regions (Ramakrishnan and Ramachandran, 1965) are associated with glycine residues. Figure 9 also shows that the majority of the residues of SGPA have phi, psi angles corresponding to those of the beta pleated sheet conformation (phi and psi are in the region of -120° and 130°, respectively).

All of the peptide bonds in SGPA, with one exception (Phe-94 to Pro-99A), are trans-peptide bonds as determined from the native electron density map. Cis-Pro-99A is located at a hairpin bend which does not have the hydrogen bonding

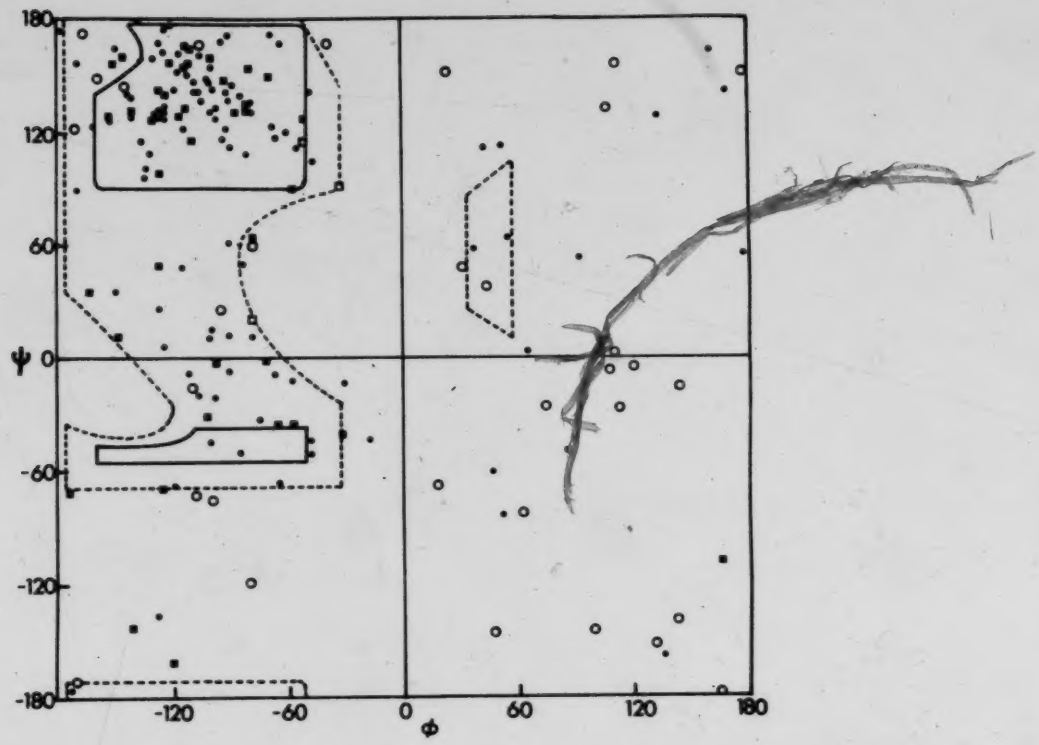


Fig. 9. Plot of the phi, psi torsional angles for the atomic model of SGPA at 2.8 angstrom resolution. The area enclosed within the solid lines of this plot is the fully allowed conformational region for tau{C(alpha)} of 110°. The broken line indicates the outer limit of acceptable van der Waals' contacts for a tau{C(alpha)} of 115°. The symbols used represent the following amino acids: (■) beta branched amino acids; (o) glycine; (□) proline; (●) other amino acid residues.

interactions of type I(10) or type II(10) beta turns (Venkatachalam, 1968). A stereo-drawing of the polypeptide chain and amino acid side chains in the vicinity of cis-Pro-99A is shown in Figure 10. Although cis-peptide linkages between amino acid residues are rare occurrences, the cis-proline peptide unit has been observed with much greater frequency in a number of globular proteins (Ramachandran and Mitra, 1976).

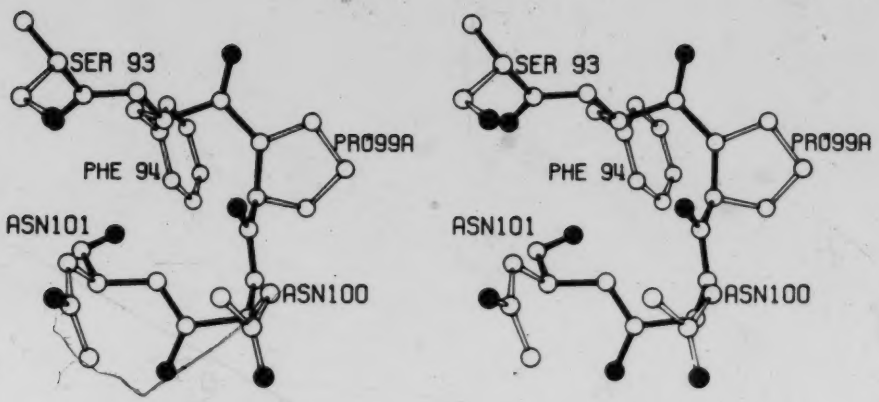


Fig. 10. Stereo-drawing of the polypeptide chain of SGPA in the region of the cis-Pro-99A peptide bond. The configuration of this bond and the accompanying conformational angles agree well with that of the reverse open turn proposed by Ramachandran and Mitra (1976).

Secondary structural features of SGPA are illustrated in the hydrogen bonding diagram of Figure 11, where hydrogen bonds between atoms of the main chain have been detailed. This Figure also shows, in a stylized fashion, the polypeptide chain folding and depicts which portions of the polypeptide chain are in sufficiently close proximity to form anti-parallel beta sheet structures. Whether a close contact was designated as a hydrogen bond or not, was based on the distance from donor to acceptor atom (less than 3.5 angstroms) and on the linearity of the putative bond. While it is not expected Figure 11 will change significantly during structural analysis of SGPA at higher resolution, the fine details of this present diagram should be regarded as tentative.

Figure 11 also indicates that the majority of the main

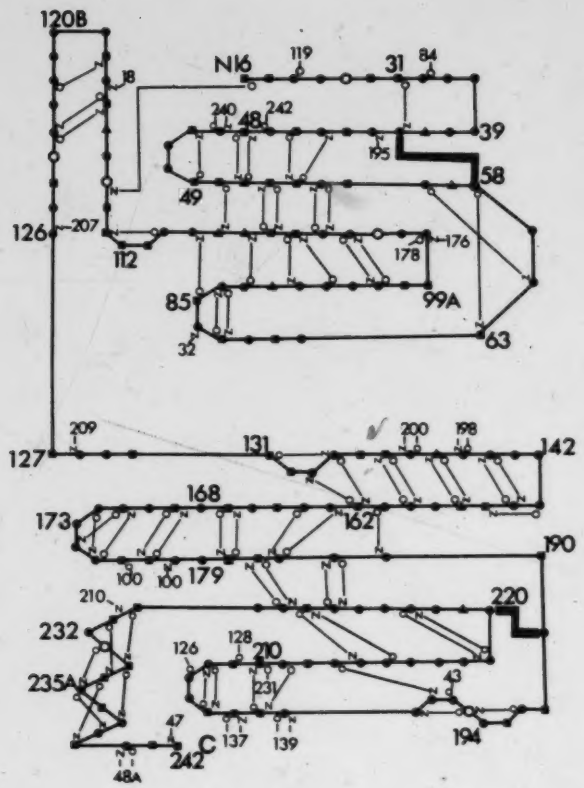


Fig. 11. Schematic drawing of the observed secondary structural features of SGPA. Hydrogen bonds between main chain carbonyl oxygen and imino nitrogen atoms are indicated. The residue numbering is that of the sequence alignment in Table 1. Symbols indicate: (o) charged acidic residues; (▲) basic residues; (●) hydrophilic uncharged residues; and (■) hydrophobic residues. The two disulfide bridges are shown as thick lines. The tertiary structural features of the main chain are shown in Figure 12.

chain hydrogen bonds are intra-domain and only a few hydrogen bonds actually link the hydrophobic cores together. This finding correlates well with similar observations for the pancreatic serine protease alpha-chymotrypsin (Birktoft and Blow, 1972). Further comparison of Figure 11 with a similar diagrammatic representation of hydrogen bonding for

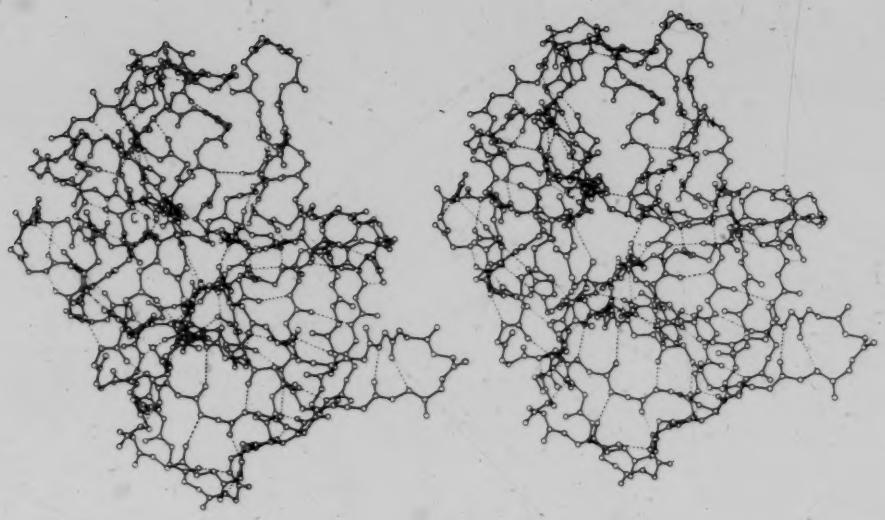


Fig. 12. Stereo-drawing of all the main chain atoms of SGPA. Hydrogen bonds shown schematically in Figure 11 are represented here by broken lines from main chain N-H groups to acceptor C=O groups. A total of 83 hydrogen bonds are presented in this drawing.

alpha-chymotrypsin (see Figure 7 of Birktoft and Blow, 1972) shows the similar polypeptide chain folding patterns, conserved in the structures of SGPA and alpha-chymotrypsin.

The same main chain hydrogen bonding pattern of Figure 11 is illustrated in the tertiary structural drawing of Figure 12. The two hydrophobic domains of SGPA are also evident in this drawing. Other hydrogen bond interactions which stabilize the tertiary structure of SGPA are listed in Table 11 (main chain to side chain) and Table 12 (side chain to side chain). These interactions were considered significant under the same criteria used for the secondary structural features discussed above. Other than the His-57 to Asp-102 interaction, there is only one salt bridge in SGPA, which is the buried one between Arg-138 and Asp-194.

TABLE 11

Main Chain to Side Chain Hydrogen Bonds

Ala-17	N	Glu-29	OE1	Asp-102	N	Gln-229	OE1
Gly-19	O	Tyr-120A	OEH	Tyr-103	O	Tyr-237	OEH
Gly-45	O	Ser-198	OG	Tyr-119	O	Ser-139	OG
Gly-56	N	Asp-102	OD2	Phe-131	N	Gln-134	OE1
His-57	N	Asp-102	OD2	Thr-142	N	Asp-194	OD2
Thr-59	O	Arg-88	NEH1	Ser-161	N	Asn-184	OD1
Ser-64	O	Trp-66	NE1	Gly-172	N	Asn-170	OD1
Ser-65A	O	Thr-33	OG1	Val-177	O	Thr-168	OG1
Ser-65A	O	Ser-64	OG	Val-190	O	Thr-226	OG1
Ile-85	O	Ser-109	OG	Gly-193	O	Ser-43	OG

TABLE 12

Side Chain to Side Chain Hydrogen Bonds

Ser-43	OG	Ser-141	OG	Thr-125	OG1	Ser-207	OG
His-57	ND1	Asp-102	OD1	Asn-129	OD1	Thr-232	OG1
Thr-59	OG1	Thr-91	OG1	Arg-138	NEH1	Thr-143	OG1
Asn-62	ND2	Thr-91	OG1	Arg-138	NEH1	Asp-194	OD1
Ser-93	OG	Tyr-103	OEH	Thr-142	OG1	Gln-192A	NE2
Asn-101	OD1	Tyr-103	OEH	Tyr-171	OEH	Ser-214	OG
Asp-102	OD1	Ser-214	OG	Asn-219	OD1	Thr-222	OG1
Thr-226	OG1	Tyr-228	OEH				

Table 13 summarizes those residues of SGPA involved in beta bends. These beta bends are characterized by the formation of a hydrogen bond from the carbonyl oxygen atom of residue 1 to the imino-nitrogen atom of the third residue, thereby forming a hydrogen bonded ring of 10 atoms. There are nine beta bends in SGPA that are either of type I(10) or type II(10). As in the case of alpha-chymotrypsin, there are a number of other hairpin turns in which the conformation does not fulfill the requirements of type I(10)

TABLE 13

Beta Bends Found in SGPA

Residues	Positions 1	2	3	4	Type
48B - 49	Val	Asn	Gly	Val	II
58 - 63	Cys	Thr	Asn	Ile	I
66 - 86	Trp	Ser	Ile	Gly	I
110 - 113	Asn	Pro	Ala	Ala	I
131 - 134	Phe	Val	Gly	Gln	II
172 - 175	Gly	Ser	Ser	Gly	I
192A - 194	Gln	Pro	Gly	Asp	II
194 - 197	Asp	Ser	Gly	Gly	II
201 - 208	Ala	Gly	Ser	Thr	I

or II(10) beta turns, but are of an intermediate nature. Residues involved in these latter turns are 33 to 40, 120A to 120D, 141 to 156 and 220 to 223. There is also the hairpin turn at cis-Pro-99A of the aspartate loop.

There is only one region in the SGPA molecule where the polypeptide chain clearly takes on a helical conformation. However, the distinction between 3(10) and alpha-helix is difficult at the present resolution. The amino acids involved in this helical region (approximately 2 turns), are located in the C-terminal region (residues 230-238). Following the two helical turns at the C-terminus, the last four residues of the polypeptide chain are in an extended conformation. A more detailed stereo-drawing of the C-terminal helical region of SGPA is given in Figure 13.

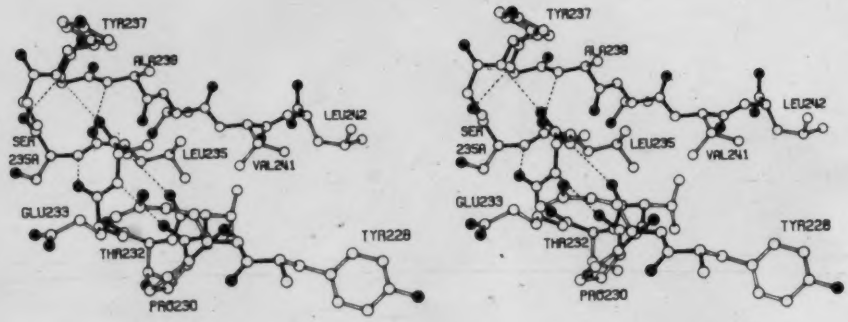


Fig. 13. Stereo-drawing of the C-terminal helical region of the SGPA molecule. Pro-230 initiates the helix; proline residues have been observed at the start of helices in a number of proteins. The final four residues of the polypeptide chain (239 to 242) are in an extended conformation. Val-241 and Leu-242 have their side chains pointing into the hydrophobic region between the two major folding domains of SGPA.

K. Structural Comparison Of SGPA and Alpha-Chymotrypsin

The alignment of the primary sequences of the bacterial and pancreatic serine proteases in Table 1, is based primarily on the results from the topological comparisons of these several serine protease structures. The topologically equivalent regions of SGPA and alpha-chymotrypsin are much more extensive than the regions of high sequence homology. Figure 14 contains the results of the topological comparison of SGPA with alpha-chymotrypsin. Figure 14a shows the overall conformation of SGPA in a stereo-drawing of the alpha-carbon atom positions connected by virtual bonds. The alpha-carbon backbone of alpha-chymotrypsin (coordinates from the Brookhaven Protein Data Bank) viewed from a similar vantage point as that of SGPA is presented in Figure 14b.

(a)

(b)

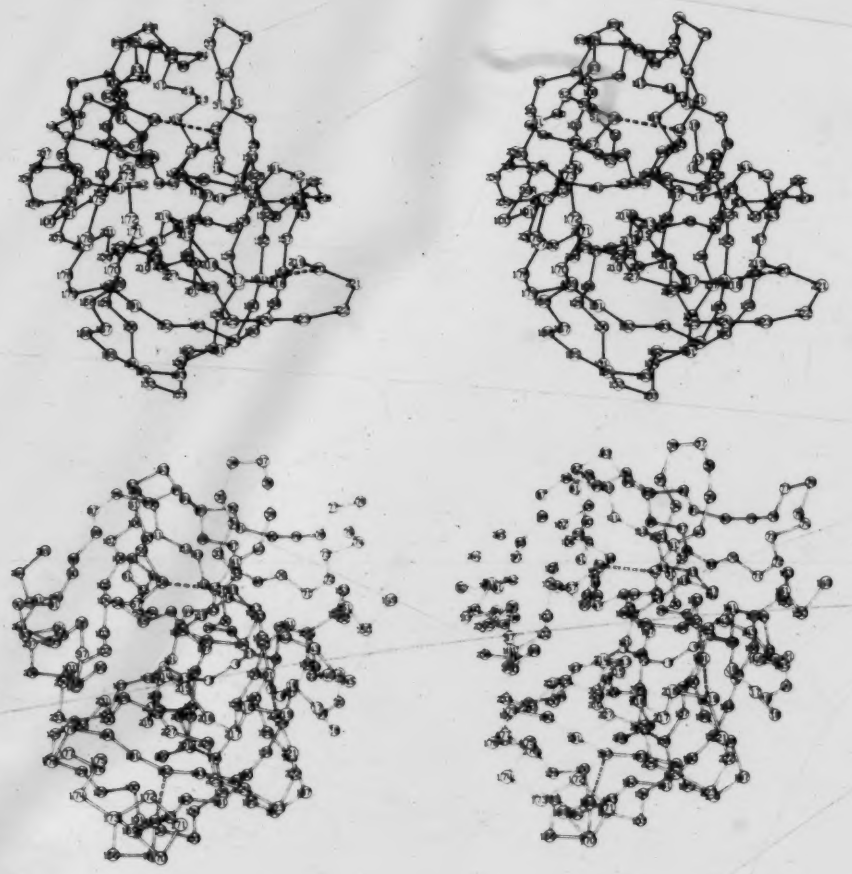


Fig. 14:
(a) An alpha-carbon backbone drawing of SGPA. The active site region is located in the center of the drawing where the alpha-carbon positions of Asp-102, His-57 and Ser-195 are evident. The two disulfide bridges 42-58 and 191-220 are denoted by broken filled virtual bonds.
(b) An alpha-carbon stereo-drawing of alpha-chymotrypsin. The view in this drawing corresponds to that view of SGPA presented above in Figure 14a. The five S-S bridges of alpha-chymotrypsin are indicated by virtual broken bonds.

Figure 15 combines the two views presented in Figure 14 and shows both enzyme structures superimposed in the orientation that resulted from maximizing their topological equivalence

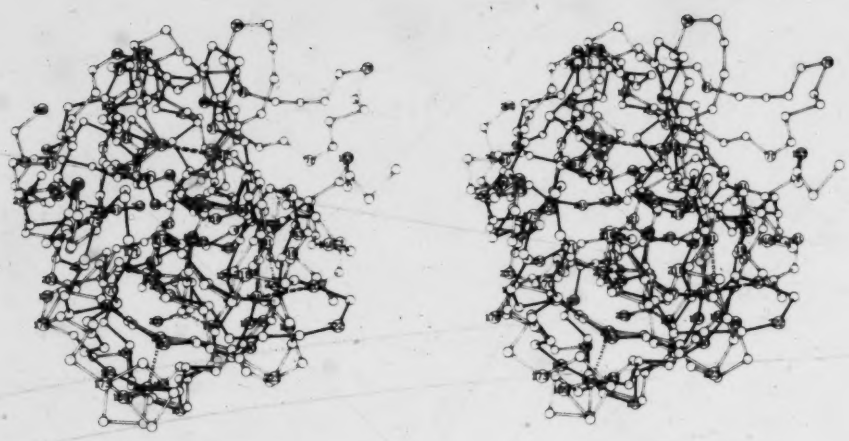


Fig. 15. An alpha-carbon drawing of SGPA (black virtual bonds) superimposed on that of alpha-chymotrypsin (open virtual bonds) to show the topological equivalence between the two enzyme structures. It is evident from this drawing that there are considerable regions of similar tertiary structure between SGPA and alpha-chymotrypsin. Also evident are important structural differences between these enzymes.

(Rossmann and Argos, 1975). As Figures 14 and 15 show, there is considerable tertiary structural homology between SGPA and alpha-chymotrypsin. In fact, there are 116 topologically equivalent residues with an r.m.s. deviation of 1.96 angstroms, if comparisons are based solely upon alpha-carbon atom positions (Table 2). The 21 chemically identical residues in the active site regions of alpha-chymotrypsin and SGPA have the most similar polypeptide chain conformations. The r.m.s. deviation of the corresponding alpha-carbon atom positions of these residues is only 0.94 angstroms.

Equally interesting and important are those regions of polypeptide chain that differ in tertiary structure between

SGPA and alpha-chymotrypsin. Alpha-chymotrypsin is synthesized as an inactive precursor, chymotrypsinogen A. Activation of chymotrypsinogen A is achieved via limited proteolysis and the formation of a free N-terminal group at Ile-16 (Hess, 1971). The completion of a salt bridge from Asp-194 to the new N-terminal group at Ile-16 is an integral part of the activation process. X-ray crystallographic studies show that zymogen activation induces movements in the polypeptide strands composed of residues 187 to 194 and 16 to 20 (Wright, 1973; Birktoft et al., 1976). These rearrangements lead to the formation of the specificity pocket and the oxyanion hole (residues 193-195, Robertus et al., 1972b).

There have been no zymogen precursors isolated for the bacterial pancreatic-like serine proteases SGPA, SGPB or alpha lytic protease. In SGPA, the crucial ion pair to Asp-194 is formed via the guanidinium group of Arg-138, an internal residue. The free N-terminal group of Ile-16, unlike in the pancreatic enzymes, does not play a role in the formation of this salt bridge. The internal nature of the Asp-194 to Arg-138 salt bridge in SGPA, indicates there is little reason to expect to find an inactive zymogen of the pancreatic type for this enzyme. Other structural features that are induced upon zymogen activation in the pancreatic serine proteases appear to be permanent features of the structure of SGPA from the time it achieves its native folded conformation.

Figure 16 illustrates the conformation of the Arg-138 to Asp-194 salt-bridge in SGPA in comparison to the Ile-16 to Asp-194 salt-bridge of alpha-chymotrypsin. It is evident from Figure 16 that the orientation of the carboxylate group of Asp-194 is similar in both SGPA and alpha-chymotrypsin, as are the positions of the guanidinium group of Arg-138 and the terminal amino group of Ile-16. The fact that this salt bridge is conserved in enzymes from bacterial sources as well as those from mammalian sources, emphasizes the important role it has in determining the enzymatically active conformation of the active site region.

The amino-terminal residue of SGPA (Ile-16) and its environment are shown in Figure 17. The terminal amino group of Ile-16 is found directed into a solvent region and is associated with a number of solvent peaks. The sec-butyl group of Ile-16 is directed partially into the N-terminal hydrophobic core. This conformation is clearly different from that observed for the N-terminal Ile-16 of alpha-chymotrypsin. This is in good agreement with the results obtained by Siegal and Awad (1973) for SGPA and SGPB, and those obtained for alpha lytic protease by Kaplan and Dugas (1969). These authors have demonstrated that N-acetylation of the N-termini of these three bacterial enzymes does not render them inactive. This is consistent with the fact that their amino-termini are not involved in the crucial salt-bridge with Asp-194. Indeed, in SGPA the terminal amino group of Ile-16 is approximately 16 angstroms

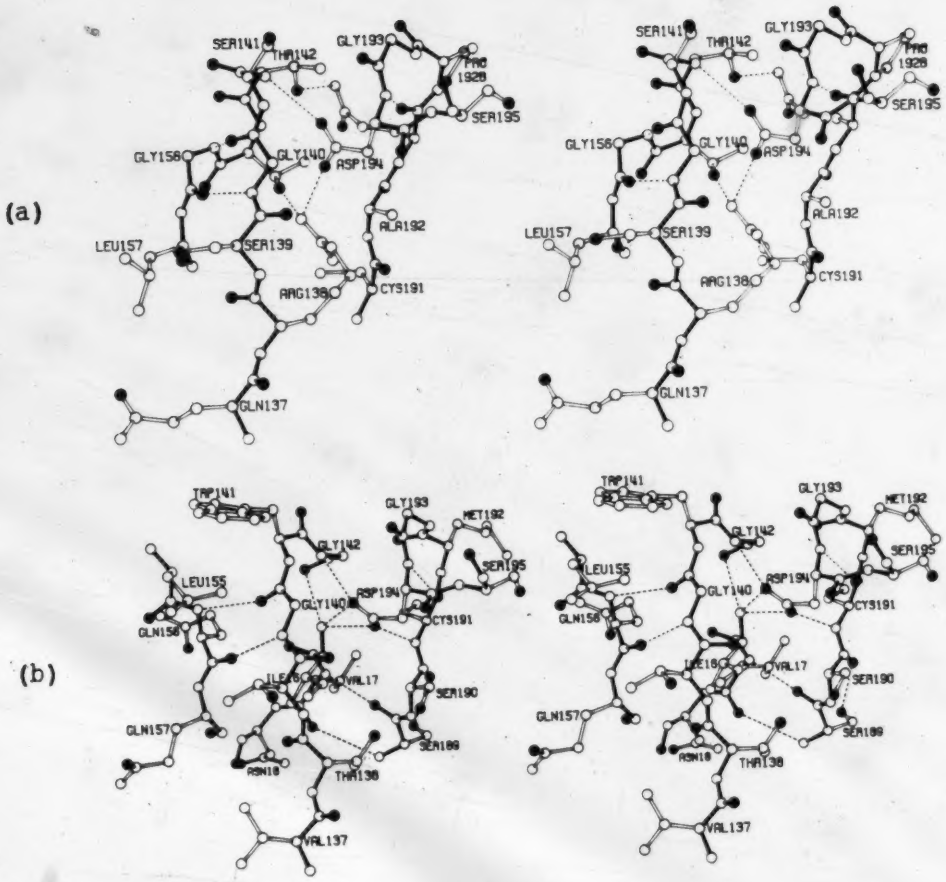


Fig. 16. A comparison of the environments of Asp-194 in (a) SGPA and (b) alpha-chymotrypsin. The SGPA molecule has a positively charged counter-ion, the guanidinium group of Arg-138, whereas in alpha-chymotrypsin the positive charge comes from the newly formed terminal amino group of Ile-16 following the zymogen activation step of this enzyme.

from the carboxyl side chain of Asp-194 (Figure 14a).

In SGPA, the polypeptide chain N-terminal to Cys-42 is 13 residues shorter than the corresponding segment in alpha-chymotrypsin (Table 1). These additional residues of alpha-chymotrypsin are required to allow the formation of the activation salt-bridge Ile-16 to Asp-194. Other than in

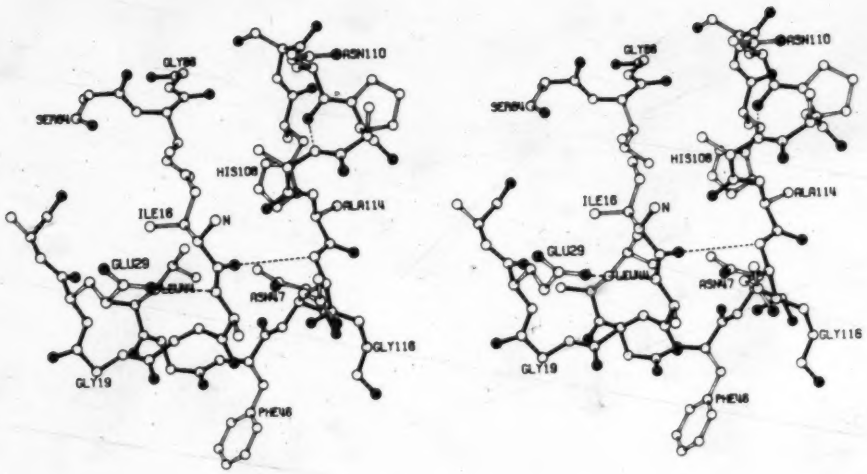


Fig. 17. The environment of the N-terminus of SGPA showing how it is accessible to acetylation without affecting the active site of this enzyme. The terminal amino group points into a solvent cavity whereas the sec-butyl side chain is directed towards the hydrophobic core of the N-terminal domain. See Figure 14a for the overall placement of this residue.

the first four residues of this portion of the polypeptide chain, there is little sequence homology between alpha-chymotrypsin and SGPA (Table 1). Even so, the sequence homology of these residues initially led to conclusions on the expected tertiary structure of this region in SGPA and other similar bacterial enzymes (McLachlan and Shotton, 1971), which are not supported in the present tertiary structure analysis. There is some structural equivalence between the two enzymes in the segment from residues 29 to 42 although there does appear to be an insertion of four residues at position 35 in alpha-chymotrypsin relative to SGPA (Figure 15).

From Cys-42, the main chain of SGPA traverses the molecule to form the histidine loop before returning to complete the disulfide bridge at Cys-58. This disulfide bridge has a very similar tertiary structure in both enzymes. It has already been suggested (Hartley et al., 1972) that the two disulfide bridges 42-58 and 191-220 are the minimum number required for an active serine protease of the pancreatic type, and the present study supports this hypothesis. The conservation of the disulfide bridge 42-58 in SGPA and all other serine proteases of the Asp-Ser-Gly structural type suggests this structural feature plays an important role in determining the proper disposition of the catalytic residue His-57, relative to the other catalytic residues of these enzymes.

The histidine loop (residues 42-58) contains four additional residues (48A to 48D) in SGPA (Table 1). Residues 48B to 49 inclusive, form a type II(10) beta bend (Table 13) at the distal end of the histidine loop. These residues occupy a volume of space which in alpha-chymotrypsin is partially taken up by the C-terminal helix (Figure 15).

In alpha-chymotrypsin, there is a large beta loop formed by residues 65 to 83 (Figure 14b). Close examination of Figure 15 shows that this loop is almost deleted in SGPA, and the polypeptide chain has a type I(10) beta turn at positions 66 to 86 (see Table 13). This loop has been termed the 'uranyl' loop because a uranyl ion is bound to elastase in this region (Shotton and Watson, 1970). In trypsin, the

uranyl loop is the binding site of a Ca^{2+} ion (Bode and Schwager, 1975) which is thought to protect the enzyme against autolysis. Additionally, the uranyl loop or Ca^{2+}-binding loop forms hydrogen bonds with the N-terminal strand near residues 29 to 34 in alpha-chymotrypsin (Birktoft and Blow, 1972). It is evident from Figure 15 that SGPA lacks an extensive Ca^{2+}-binding loop and there are a number of major structural differences between SGPA and alpha-chymotrypsin in this region.

The fourth major loop in the N-terminal domain of SGPA is the aspartate loop (residues 87-108). This loop is so-named because it contains the catalytically important, buried Asp-102. The aspartate loop of alpha-chymotrypsin is five residues longer than the equivalent aspartate loop of SGPA. Reference to Figure 15 shows that these additional residues in alpha-chymotrypsin are inserted at the cis-Pro-99A bend of the corresponding loop in SGPA. Residues 95 to 99 in alpha-chymotrypsin isolate the Asp-102 to His-57 interaction from direct solvent contact. It is evident from an examination of Figure 15 that the residues occupying the corresponding region in SGPA come from the methionine loop (residues 164-182). The positions of residues 174 to 177 and of other neighbouring residues, have also left the side chain of Asp-102 in SGPA relatively inaccessible to solvent. In SGPA, as in alpha-chymotrypsin, the aromatic ring of Phe-94 is on the surface of the molecule partially isolating Asp-102 from direct contact with the solvent media. Other

segments of the aspartate loop, residues in the beta bend of the methionine loop and the main polypeptide chain near His-57 serve to complete this barrier to solvent in SGPA (Figure 19).

In spite of the very different polypeptide chain conformations in the vicinity of Asp-102 in SGPA relative to alpha-chymotrypsin, the orientation of the beta-COO^- group of this residue is essentially the same in both enzymes (Figure 19). In addition, the strand of main chain from residues 101 to 108 has a very similar conformation in these two enzymes (Figure 15).

With the completion of the aspartate loop at residue 108, the polypeptide chain extends into the C-terminal folding unit of SGPA and initiates the second major hydrophobic domain. A number of structural changes are evident in the region of main chain from residues 113 to 191 when one compares the tertiary structure of SGPA to that of alpha-chymotrypsin. (Figure 15).

The first such conformational difference is at position 117, where a small beta loop (residues 117 to 124) is formed at the back of SGPA (Figure 14a). This loop has no counterpart in alpha-chymotrypsin where the equivalent volume is occupied by residues 1 to 6 and 23 to 28. This relatively small conformational difference seems to be associated with deletions in the N-terminal strand (residues 20-28) and the uranyl loop (residues 67-83) of SGPA relative to alpha-chymotrypsin (Table 1). There is also a four

residue insertion in SGPA at position 120 which contributes to these structural changes.

The first major beta loop of the C-terminal hydrophobic domain in SGPA is topologically equivalent to the autolysis loop of alpha-chymotrypsin (Figure 14). The amino acid residues involved in this prominent structural feature of both molecules, extend from position 131 to position 163. In earlier sequence alignments, this loop in the bacterial serine proteases has been referred to as the methionine loop. However, the recent realignment of the pancreatic and bacterial enzymes, based on topological equivalences (Table 1) has shown that it is unnecessary to name this region of the molecule differently from that in the pancreatic enzymes. Even though there is a great deal of tertiary structural homology (i.e. strands 131-142 and 156-164), there is very little sequence homology between SGPA and alpha-chymotrypsin in this region. Only three residues of the 21 topologically equivalent residues have identical sequences (Table 1). It would appear that only one of these three identical residues is structurally required, Gly-140. A side chain at position 140, even as small as the methyl group of alanine, would sterically interfere with the side chain of the active site residue Asp-194.

As referred to above, the positive charge which neutralizes the buried carboxyl group of Asp-194 comes from the guanidinium group of Arg-138. This basic residue is common in all three of the bacterial serine enzymes, SGPA,

SGPB and alpha lytic protease (Table 1). The guanidinium group of Arg-138 is in a position that is structurally homologous to the position of the charged amino terminus of the activated pancreatic enzymes (Figure 16), and its presence implies that there is no zymogen for these three enzymes.

The major structural differences between the autolysis loop of SGPA and those of alpha-chymotrypsin, trypsin and elastase are the insertion of 12 amino acids in the sequences of the pancreatic enzymes at residue 144 and the replacement of the arginine residue at position 138. It is these polypeptide chain differences that form the necessary open type structure allowing the N-terminus of the activated pancreatic enzymes to form the crucial buried salt-bridge to Asp-194.

The methionine loop of SGPA (residues 164 to 182) has a completely different conformation from that of the pancreatic enzymes (Figure 15). In SGPA these 19 residues form a large beta loop, the bend of which is near the aspartate loop and, along with Phe-94, serves to isolate the Asp-102 to His-57 interaction from solvent. This loop in alpha-chymotrypsin is more compact and is situated at the lower extremity (in Figure 14b) of the enzyme. Assuming substrate binding occurs as postulated for the pancreatic enzymes (Segal et al., 1971; Ruhlmann et al., 1973; Sweet et al., 1974), the methionine loop of SGPA is positioned close to the expected secondary substrate binding region of this

enzyme. Thus, it is reasonable to expect from the prominent positioning of the methionine loop in the active site of SGPA, that it plays an important role in forming secondary binding sites for SGPA. Recent kinetic studies of SGPA with several synthetic substrates (Bauer et al., 1976a,1978) have confirmed the presence of several important secondary binding subsites for this enzyme.

Figure 15 shows that from residues 179 to 184 there is considerable topological equivalence between SGPA and alpha-chymotrypsin. Following this strand, there is a small loop 185-189 present in alpha-chymotrypsin (Table 1). The absence of these residues from the sequence of SGPA, coupled with a two residue insertion at Ala-192 and the presence of the side chains of Val-190, Thr-226 and Tyr-228, makes the primary specificity pocket of this enzyme more of a shallow surface depression than in alpha-chymotrypsin.

The alpha-carbon atom of Cys-191 in SGPA is in a topologically equivalent position to that of Ser-189 in alpha-chymotrypsin. Rather than renumber the disulfide bridge (191 to 220), which is highly conserved in the serine proteases of the Asp-Ser-Gly type, this has been considered a conformational alteration occurring during the evolution of the pancreatic-type structure and the original numbering has been retained (Table 1). The differing conformation of this disulfide bridge appears to result from two residues (192A and 192B) being deleted on going from the bacterial structure to the pancreatic structure. Despite the different

disulfide bridge conformations of SGPA and alpha-chymotrypsin, the nearby active site residues Asp-194, Ser-195 and those forming the oxyanion hole, retain very similar conformations in both enzymes.

The overall conformation of the polypeptide chain of residues 192 to 197, which contains two beta bends (Table 13), is highly conserved both in tertiary structure (Figure 15) and in primary structure (Table 1) in SGPA and the pancreatic serine proteases. The two beta bends formed in this region are anchored by the disulfide bridge 191 to 220, and the ion pair to the carboxyl group of Asp-194.

Beginning at Ser-195, the polypeptide chain forms a beta loop by traversing the central portion of the C-terminal domain and bending back upon itself to return to the active site region at Ser-214. Although the serine loop of SGPA is four residues shorter than in alpha-chymotrypsin, the overall conformation of this loop in both enzymes is similar (Figure 15). The four residue deletion in SGPA occurs at the distal end of the loop (residues 203-206) and has no affect on the orientation of active site residues.

Residues 214 to 228 form a large beta loop (specificity pocket loop) which constitutes one side and part of the bottom of the primary specificity site, S1. It has been shown that in gamma-chymotrypsin (Segal et al., 1971) and in the trypsin-bovine pancreatic trypsin inhibitor complex (Ruhlmann et al., 1973), residues 214 to 217 are involved in an anti-parallel beta structure with the peptide chain of

the bound inhibitors studied. These residues are homologous in primary structure and also in tertiary structure with those of SGPA, suggesting that a substrate binding mode similar to that of gamma-chymotrypsin or trypsin is also possible for SGPA. Further residues of the specificity loop (218 to 225) have somewhat different conformations in SGPA and alpha-chymotrypsin, resulting from the structural differences present at the disulfide bridge 191 to 220.

The C-terminal helical segment of the polypeptide chain in SGPA starts at residue Pro-230. There are only two turns of helix which appear to be a mixture of 3(10) and alpha-helix (see Figure 13). This helical region is topologically equivalent in SGPA and alpha-chymotrypsin (Figure 15), but is shorter in SGPA. The last four residues of SGPA are not in a helical conformation but are in an extended conformation lying on the enzyme surface. This feature of SGPA contrasts sharply with the C-terminal helical conformation observed for all three of the pancreatic serine proteases (Birktoft and Blow, 1972; Sawyer et al., 1978; Bode and Schwager, 1975), in which the C-terminal helix extends all the way to the C-terminus. The C-terminal carboxyl group of Leu-242 in SGPA does not make contacts with other groups of the enzyme, but is associated with peaks of electron density representing bound solvent molecules. The hydrophobic side chain of Leu-242 points into the hydrophobic contact region between the two folding domains of SGPA.

L. SGPA As An Evolutionary Precursor Of Alpha-Chymotrypsin

In comparisons of SGPA with alpha-chymotrypsin (Figure 14 and 15) one is struck by the conservation of tertiary structure (64% topological equivalence, Table 2) in light of the absence of significant seguence homology (21% sequence identity, Table 1). As can be seen in Figure 14 and 15, the most highly conserved tertiary structure is located about the catalytic residues Asp-102, His-57 and Ser-195; also the site of the most highly conserved polypeptide sequences. However, many stretches of polypeptide chain with little sequence homology are nevertheless conserved in tertiary structure.

Close examination of the tertiary structures of SGPA and alpha-chymotrypsin shows that, aside from minor changes in surface loops, there are only two major tertiary structural differences between these enzymes. One of these is related to the presence or absence of a zymogen function; the other being connected with rearrangements in the substrate binding region.

The structural changes that would be required for the bacterial structure to incorporate a zymogen precursor function are four-fold. Firstly, the N-terminal loop (residues 16-42) must be increased in size by 13 residues in order to accommodate the proper conformation of the N-terminus to be involved in an ion-pair with Asp-194. A second change requires increasing the size of the autolysis

loop of SGPA by 12 residues (residues 144-155). These additional residues are required to provide the appropriate surrounding structure in which the Asp-194 to Ile-16 salt bridge can be completed. Thirdly, a specific zymogen precursor polypeptide must be added to the N-terminal of SGPA, which is cleaved at the appropriate time and site for enzymatic activation to occur. Finally, the uranyl loop (residues 65-84) must also be increased in size and form stabilizing contacts with a now longer N-terminal portion of the polypeptide chain.

The evolution of SGPA into alpha-chymotrypsin in the substrate specificity region, would also require a number of structural alterations. The methionine loop (residues 164-182) would have to take on a more compact conformation and be tucked down to the lower extremity of SGPA. In order to compensate for the loss of protection the methionine loop had afforded Asp-102, it would be necessary to increase the size of the aspartate loop by five residues. Further alterations would be necessary to increase the size of the S1 binding site so that it is as large as that of alpha-chymotrypsin. These include the deletion of residues 192A to 192B and the insertion of four additional residues at position 184.

Thus a number of significant insertions and deletions would be required to transform SGPA into a model of alpha-chymotrypsin. Nevertheless, it is not unrealistic to suggest SGPA is a model for the evolutionary precursor of

alpha-chymotrypsin, especially in light of the preservation of very similar catalytic centers and tertiary structural cores. Development of zymogen control and greater cleavage specificity in mammalian alpha-chymotrypsin very probably arose from the need for greater biological control of such serine proteases when incorporated in the processes of a more sophisticated organism.

M. Active Site Region

Figure 18 shows several sections of the native electron density map of SGPA in the region of the active site. The protein-solvent boundary on this map is clearly evident. As can be seen in this Figure, there is ordered electron density in close proximity to the side chains of Ser-195 and His-57 which we have tentatively interpreted as partially occupied phosphate ions or solvent molecules. Figure 18 shows that the side chain of Phe-94 lies just above the carboxylate of Asp-102. Also shown, is the hydroxyl group of the side chain of Tyr-171, which interacts with the side chain of Ser-214. It is evident that the gamma oxygen atom of this latter residue is in close proximity to the carboxylate group of Asp-102. The well ordered side chains of Thr-59 and Asn-62 are visible to the left of the density associated with the imidazole ring of His-57. In the background, close to His-57, the electron density associated with the disulfide bridge 42-58 can be discerned.

In Figure 19, SGPA and tosyl alpha-chymotrypsin (the

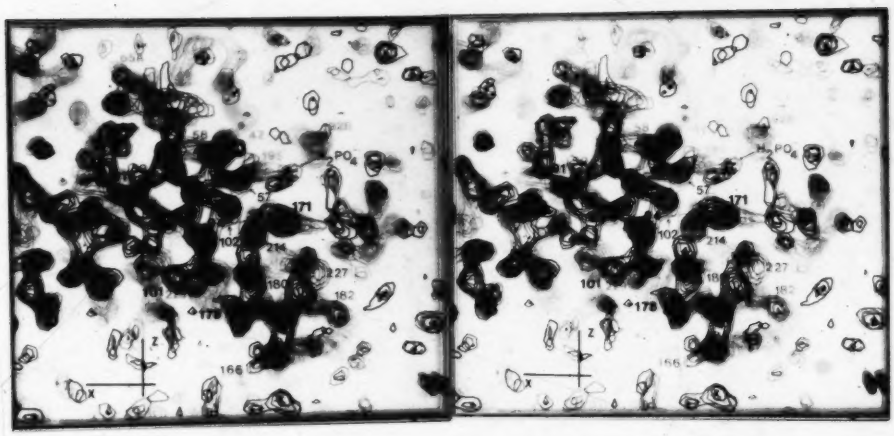


Fig. 18. A stereo-representation of the multiple isomorphous replacement phased 2.8 angstrom resolution electron density map of SGPA. Shown above is the region of the molecule containing the active site. The first contour is drawn at 0.56e/(angstroms)^3 (including 0.23e/(angstroms)^3 contributed by the F(000)/V term) and subsequent contours are drawn at progressive intervals of +0.11e/(angstroms)^3. The view of the map presented here looks directly down the b axis of the unit cell and includes a cross-section of one complete molecule of SGPA. Four residues in the active site His-57, Asp-102, Ser-195 and Ser-214 have been labelled and are situated in the central portion of the map. Also shown are the alpha-carbon positions and sequence numbers (Table 1) of some of the other residues which are evident on these sections of the electron density map.

tosyl group has been omitted) have been compared in the regions of their respective active sites. The same orientation is presented in this Figure as had been used in Figure 14 and 15. Not only are the main chain active site conformations very similar in the two enzymes but also many of the side chains have similar orientations, e.g. His-57, Asp-102, Ser-214, Cys-42 and Cys-58. These similar tertiary structural features combined with the detailed kinetic measurements of the hydrolysis of polypeptide substrates (Bauer et al., 1976a,b) indicate that the catalytic

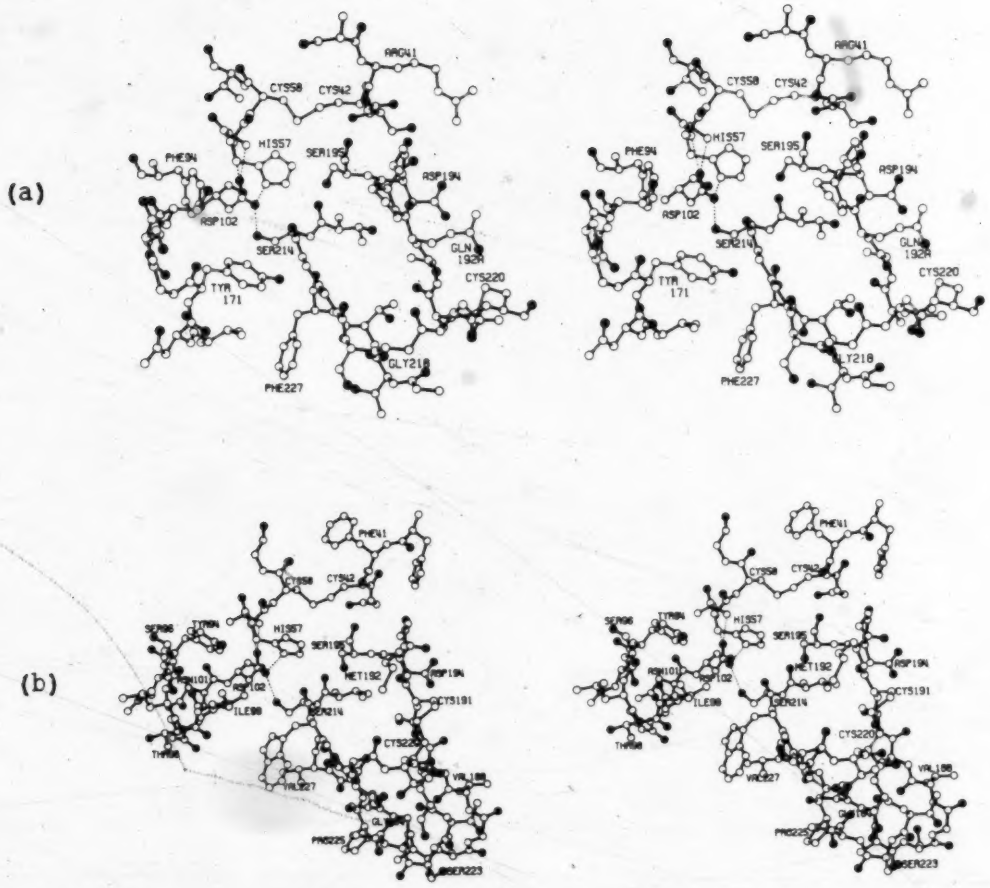


Fig. 19. Stereo-drawing of the active site region of (a) SGPA, and (b) alpha-chymotrypsin, in a similar orientation. The polypeptide main chain bonding is shown with solid black bonds and oxygen atoms are distinguished by solid black circles. Only hydrogen bonds between active site residues and those with surrounding polypeptide chain have been illustrated as broken lines. The alpha-chymotrypsin coordinates used were those of the tosylated enzyme although the tosyl group has been omitted in this drawing for clarity.

mechanism for the pancreatic and the microbial

pancreatic-like serine proteases is similar.

In SGPA a number of hydrogen bonds are formed among the

residues in the active site and with surrounding polypeptide chains. However, the interaction between NE2 of His-57 and CG of Ser-195 is probably only a weak hydrogen bond, since the geometry of this bond is significantly distorted from an ideal conformation. This observation has also been made for other serine protease structures (Kraut, 1977; Matthews et al., 1977). The active site residue to which the greatest number of hydrogen bonds are made is the buried carboxylate of Asp-102. This group is the recipient of four hydrogen bonds and as such is in a hydrophilic environment. It is because the carboxylate of Asp-102 is in this polar environment, albeit segregated from the surrounding solvent medium, that it is likely the carboxyl group of this residue has a pKa in the normal range.

The present study of SGPA confirms that this enzyme retains a similar conformation of active site residues (Asp-102, His-57, Ser-195) as found in the pancreatic serine proteases and also implicates the side chain of Ser-214 as a important component in the active site. Ser-214 is a highly conserved residue in the pancreatic family of serine proteases, and an analog is also found in the structurally distinct subtilisin enzyme (Ser-33 is hydrogen bonded to the carboxylate side chain of Asp-32; Kraut et al., 1971). In all of the pancreatic enzymes and SGPA, the hydroxyl group of Ser-214 participates as a hydrogen bond donor to one of the carboxyl oxygen atoms of Asp-102 (Figure 19). Although it is unlikely that the O-H proton is involved directly in

the catalytic process, one expects that this hydrogen bond, in conjunction with the two peptide N-H hydrogen bonds to Asp-102 (from residues Gly-56 and His-57), would stabilize the negative charge on the carboxylate group of Asp-102. It is of interest to note that in the subtilisin structure the serine residue which has this hydrogen bond donor role is not topologically equivalent to Ser-214 of the pancreatic enzymes (Kraut et al., 1971). Nevertheless, it should be stressed that the carboxylate oxygen atom of the active site aspartate residue (Asp-32 in subtilisin, Asp-102 in SGPA and the pancreatic enzymes), which is hydrogen bonded to the active site histidine residue, is also simultaneously hydrogen bonded to the gamma oxygen atom of a second serine residue (Ser-33 in subtilisin, Ser-214 in SGPA and the pancreatic enzymes).

Figures 15 and 19 show that the polypeptide chain from residues 191 to 196 has an almost identical conformation in SGPA and alpha-chymotrypsin, in spite of sequence differences in the region of residues 192-193. This segment of polypeptide chain performs an important role in providing the correct conformation to produce the oxyanion hole (Robertus et al., 1972b). Because the oxyanion hole is so similar in SGPA and alpha-chymotrypsin, it is also reasonable to expect a similar orientation and polarization of the carbonyl bond of a susceptible peptide bond in both enzymes. The zymogen activation phenomenon already alluded to previously, is reliant on the formation of the salt

bridge to Asp-194 and this mechanism in the pancreatic enzymes provides the active conformation of the oxyanion hole which is a permanent feature in SGPA.

Examination of Figure 19 shows that SGPA does not have the well formed 'tosyl hole' primary binding site as found in alpha-chymotrypsin. The primary specificity site of SGPA is a shallow groove formed from three stretches of polypeptide chain including residues 191 to 192B, 213 to 219 and 224 to 227. For the most part the walls of this binding site comprise the peptide planes linking these residues. One feature of the specificity pocket region, which is highly conserved in both SGPA and alpha-chymotrypsin, is the path of the polypeptide chain of residues 213 to 217. This strongly suggests that SGPA binds substrate polypeptide chain backbones in a manner analogous to the pancreatic enzymes via an anti-parallel beta structure (Segal et al., 1971).

A major factor in the poorly defined nature of the specificity pocket of SGPA appears to be a two residue insertion at Ala-192, which results in the rearrangement of the disulfide bridge between residues 191 to 220. This altered disulfide bridge conformation causes residues 218 to 220 to turn sharply inwards, narrowing the opening to the S1 binding region, and bringing residues 224 to 227 directly under the specificity pocket region (Figure 19). These changes in the main chain conformation and the associated resulting placement of the side chain of Thr-226, makes the

S1 binding cavity of SGPA rather shallow relative to that of alpha-chymotrypsin.

Solution substrate studies carried out on both SGPA and alpha-chymotrypsin also reflect the structural differences observed between the two enzymes (Bauer et al., 1976a,b). These studies involved the hydrolysis of specifically designed tetrapeptide substrates and indicated that the binding pocket of SGPA is less developed and therefore less specific towards substrate side groups. Thus, peptides with P1 side chains as small as alanine and as large as tyrosine are hydrolyzed. Nevertheless, residues with the larger P1 hydrophobic side chains were preferred. These results correlate very well with model building studies done with SGPA which indicate that tyrosine is the largest amino acid to fit easily into the S1 binding pocket and that tryptophan cannot be accommodated without forming prohibitively close contacts. Somewhat different results were observed in the corresponding substrate studies with alpha-chymotrypsin which show that only the side chains of phenylalanine, tyrosine and tryptophan bind with appreciable affinity for the S1 specificity pocket. This is undoubtably a reflection of the more developed and specific nature of this pocket in alpha-chymotrypsin.

Additional substrate kinetic studies with SGPA (Bauer et al., 1976a,b; Bauer, 1978) indicate that enhanced enzymatic activity and protein-substrate interactions occur with longer substrates. It is particularly easy to locate

the most probable S2 site of SGPA since a large hydrophobic pocket lies in the region expected to bind this residue. Here it is assumed that the substrate binds to the enzyme in an anti-parallel beta sheet conformation as it does in gamma-chymotrypsin (Segal et al., 1971). This pocket in SGPA is formed by Tyr-171, His-57, Phe-94 and the polypeptide chain residues 172 to 175. It is probable that for SGPA, secondary interactions are a necessary prerequisite for proper substrate orientation as a result of less specific binding in the S1 specificity site.

N. Conformation Of Ser-195

SGPA has been crystallized in the native state thereby obviating difficulties in the interpretation of the position of the side chain of Ser-195, a problem encountered in other serine protease structural studies, where this residue has been derivatized (Kraut, 1977). The side chain of Ser-195 in SGPA has a torsional angle, chi1, of approximately -80° from our interpretation of the 2.8 angstrom resolution multiple isomorphous replacement map. This conformation is similar to those reported for chymotrypsinogen (-60°, Birktoft et al., 1976), elastase (-84°, Sawyer et al., 1978), native bovine trypsin (-81°) and the trypsin-pancreatic trypsin inhibitor complex (-83°) (Bode et al., 1976).

In SGPA, the contact distance from NE2 of His-57 to OG of Ser-195 is quite short (2.7 angstroms); however, it can be seen in Figure 19 that OG of Ser-195 is out of the plane

of the imidazole ring of His-57. It is in fact approximately 2.0 angstroms from the ideal hydrogen bonding position to NE2. Therefore, this interaction, albeit apparently short, represents a significantly distorted hydrogen bond. This finding supports the suggestion recently made by Kraut (1977) that the hydrogen bond between OG of the reactive serine and NE2 of the histidine is distorted to the point where it is either non-existent or at best very weak.

The proposed 'up' position for OG of Ser-195 in native alpha-chymotrypsin (Birktoft and Blow, 1972; Blow, 1976) would appear to be an artifact resulting from the derivation of the native structure from a difference Fourier electron density map (Henderson, 1970; Steitz et al., 1969). This 'up' position (chi1=90°) has not been observed in SGPA, nor in the very highly refined structure of uncomplexed trypsin (Bode and Schwager, 1975) or in any other serine protease structure. However, as shown in Figure 19, the conformation of Ser-195 in tosyl alpha-chymotrypsin is more like that of SGPA and other serine proteases.

IV. The Tertiary Structure Of *Streptomyces griseus* Protease B At 2.8 Angstrom Resolution

A. Structure Determination

SGPB was isolated from pronase (Jurasek *et al.*, 1971) by ion-exchange chromatography on CM-sephadex and generously provided for these experiments by Drs. L. Jurasek and L.B. Smillie. Three crystalline modifications of SGPB have been reported (Codding *et al.*, 1974). The 2.8 angstrom resolution crystal structure discussed herein is that of the orthorhombic modification grown from 0.7M KH(2)PO(4) at pH 4.2. The procedures of crystallization, data processing and completing a preliminary chain tracing for this crystalline modification of SGPB have been described (Delbaere *et al.*, 1975). These procedures were similar to those described in the elucidation of the structure of SGPA. This preliminary report was the first to describe the overall polypeptide chain folding of the pancreatic-like microbial serine proteases.

B. Interpretation Of The Electron Density Map

A native electron density map of SGPB was computed with the native structure factor amplitudes (having I > 3sigma(I)) and the best phases derived from the final phasing cycle. The detailed conformation of the polypeptide chain of SGPB was determined from a native electron density map plotted on a scale of 2cm/angstrom. The grid of this map

was 0.75 X 0.75 angstroms, with sections of electron density computed 0.75 angstroms apart perpendicular to the c axis.

The first contour of the native electron density was drawn at 0.35e/(angstroms)^3 (including the F(000)/V term of 0.22e/(angstroms)^3) with subsequent intervals increasing by 0.13e/(angstroms)^3. The standard error of this native electron density map was determined as 0.178e/(angstroms)^3 (Cruickshank, 1949; Dickerson et al., 1961). Contour lines were traced directly onto cellulose acetate sheets for use in a Richards optical comparator (Richards, 1968). A Watson-Kendrew model was then constructed in the usual manner. In the model fitting procedure Ala-84 was reinterpreted as Trp-84 (sequence numbering of Table 1) in an ambiguous region of the published sequence (Jurasek et al., 1974). Also, a valine residue was inserted at position 177 (new numbering) to be consistent with the interpretation of the electron density map in this region.

The positions of all non-hydrogen atoms of SGPB were measured from the resultant Watson-Kendrew model using the plumb-line method. Following a similar procedure employed for SGPA, these coordinates were used in Diamond's Class II model building procedure (Diamond, 1966,1974) as guide coordinates to achieve the optimal fit to a stereo-chemically correct structure. The overall r.m.s. deviation between the measured and the model built coordinates was 0.25 angstroms.

C. Molecular Conformation And Comparison With SGPA And Alpha-Chymotrypsin

The overall shape of SGPB is roughly globular, with outer dimensions of approximately 44 X 40 X 27 angstroms. Like SGPA, the polypeptide chain of SGPB is folded such that there are two structurally similar hydrophobic domains. The juncture of these domains forms a shallow surface depression which contains the active site region. The two hydrophobic cores are each composed of six beta strands hydrogen bonded to produce a beta barrel structure (Birktoft and Blow, 1972). This polypeptide chain conformation can be described as +1, +1, +3, -1, -1 for both domains, in the notation of Richardson (1976). Figure 20 schematically depicts the hydrophobic domain folding of SGPB. This figure also accurately describes the polypeptide chain folding found in SGPA. As in SGPA, the four beta loops of SGPB which form the amino-terminal domain are, the N-terminal, histidine, uranyl and aspartate loops. Polypeptide chain loops of the carboxy-terminal domain are, the autolysis, methionine, serine and specificity loops.

All of the peptide bonds of SGPB, with one exception (Phe-94 to Pro-99A), were found to be trans-peptide bonds. Cis-Pro-99A is located at the hairpin beta turn of the aspartate loop (residues 87-109), in a conformation very similar to that found for this same residue in SGPA. This cis-peptide unit appears to be a common feature of the pancreatic-like class of microbial serine proteases. Like

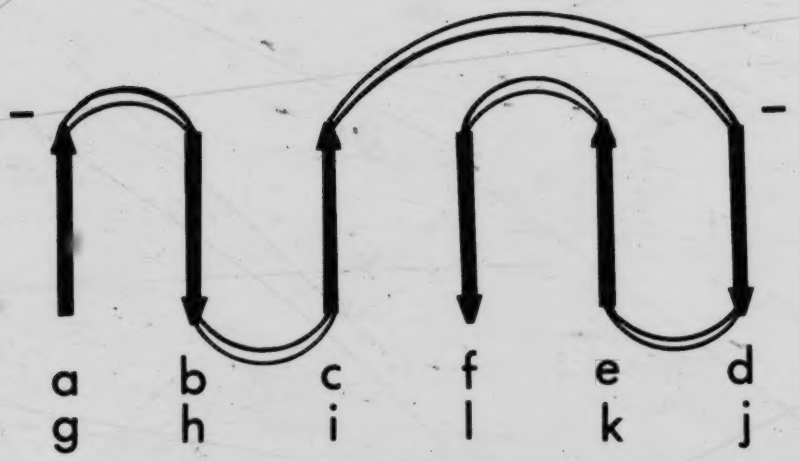


Fig. 20. Diagrammatic representation of the two six stranded beta barrels of SGPB. Strands of the polypeptide chain are labelled a-f (barrel 1) and g-l (barrel 2) from the N to the C terminus. The residue numbers at the start and end of each strand for the first beta barrel are: a(18-33), b(39-48B), c(48D-55), d(60-81), e(83-94) and f(101-107); for the second beta barrel: g(132-140), h(143-163), i(181-190), j(195-201), k(208-219) and l(223-230).

SGPA, the majority of residues in SGPB are in a beta sheet conformation. There is only one region of the SGPB molecule in which the polypeptide chain takes on a clearly defined helical conformation; that being near the C-terminal end of the enzyme.

There are a number of salt bridges formed in the structure of SGPB, in addition to the structurally important Arg-138 to Asp-194 interaction, which is also present in SGPA. Interestingly enough, both polypeptide chain termini are involved in such interactions. Additional salt bridges present in SGPB, but not in SGPA, include: Ile-16 to Asp-116; Asp-29 to Arg-139; Arg-48A to Tyr-242(OT1) and Lys-115 to Tyr-242(OT2).

The path of the alpha-carbon backbone of SGPB is illustrated in Figure 21a, in a view looking into the active site region. Figure 21a has the same view as similar drawings of SGPA and alpha-chymotrypsin (Figure 14). A topological comparison of the polypeptide chain folding of SGPB with that of SGPA and alpha-chymotrypsin is given in Figures 21b and 22 respectively. It is not unexpected that SGPA and SGPB are structurally similar. These proteases are not only synthesized by the same bacterial organism but also have 61% identity of primary sequence (Table 2). Further comparison of the tertiary structures of these two enzymes shows they have 85% topological equivalence within a r.m.s. deviation of 1.46 angstroms. This remarkable degree of structural homology between the structures of SGPA and SGPB is clearly evident in Figure 21b.

Comparison of the tertiary structures of SGPA and SGPB, shows there are only three areas of significantly different conformations between these two enzymes. All of these are restricted to surface polypeptide loops. One such conformational change can be seen in the region of the uranyl loop (residues 65A-86). While this loop is virtually absent in SGPA; SGPB has a seven residue insertion in its polypeptide sequence at this point (Table 1). As a consequence, a modestly sized uranyl loop is present in the structure of SGPB (Figure 21b). A further difference between SGPA and SGPB resides in the small beta loop formed to the back of each enzyme (residues 117-124 in SGPA). The primary

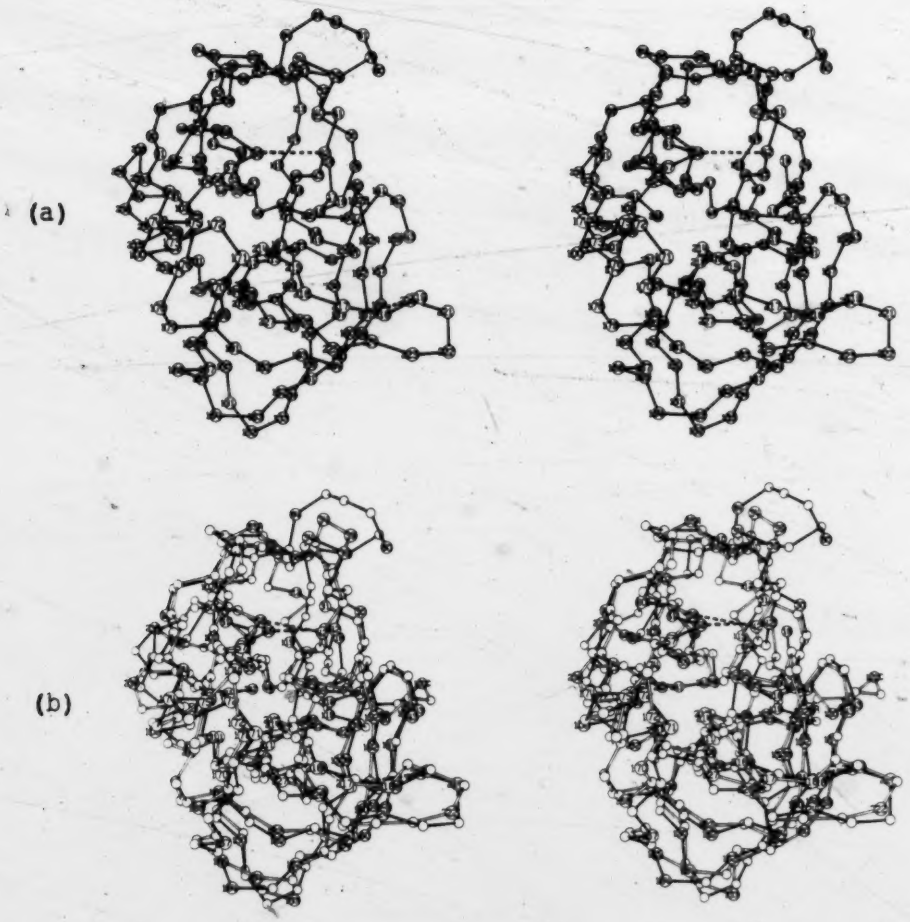


Fig. 21:
(a) Stereo-drawing of the polypeptide chain folding of SGPB in a view down the c axis of the unit cell; only alpha-carbon atom positions are indicated. Disulfide bridges are shown as dashed virtual bonds.
(b) A stereo-view illustrating the topological comparison of the polypeptide chains of SGPB (dark bonds) and SGPA (open bonds); there are 154 residues topologically equivalent within a r.m.s. deviation of 1.46 angstroms in these two enzymes. Disulfide bridges present in each enzyme are depicted as dashed virtual bonds.

sequence of SGPB has four fewer residues in this region of the polypeptide chain, resulting in the formation of a much smaller beta loop in this enzyme. Finally, residues 173 to 174 have a somewhat different conformation in the two

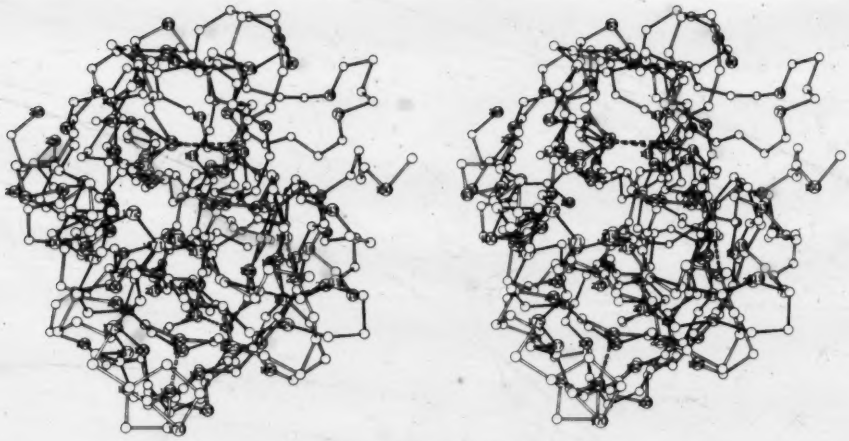


Fig. 22. A topological comparison of the polypeptide chain of SGPB (dark virtual bonds) and alpha-chymotrypsin (open virtual bonds) viewed in stereo, down the crystallographic c axis of SGPB. The circles represent alpha-carbon atom positions. Disulfide bridges are depicted as dashed virtual bonds between the corresponding alpha-carbon atoms. There are 117 residues, in the two enzymes, which are topologically equivalent within an r.m.s. deviation of 2.07 angstroms.

enzymes. However, this structural difference may be caused by crystal packing forces, since in crystals of SGPB, an intermolecular ion pair is formed between the side chain of Asp-175 and that of Arg-48A of a crystallographically related enzyme molecule (x,y,z-1).

Alignment of the primary sequence of SGPB with that of alpha-chymotrypsin indicates there is little primary sequence homology between these enzymes (18%, Table 2). Nevertheless, structural comparisons of these enzymes, in an orientation resulting from maximizing their topological equivalence (Figure 22), demonstrates the considerable tertiary structure homology shared by SGPB and

alpha-chymotrypsin. Indeed, there are 117 topologically equivalent residues (63% of the residues of SGPB) with a r.m.s. deviation of 2.07 angstroms, when comparisons are based solely on alpha-carbon positions (Table 2).

The observed structural homology between SGPB and alpha-chymotrypsin is not surprising in light of the close structural relationship between SGPA and the mammalian enzyme (Figure 14), and the remarkably similar tertiary structures of SGPA and SGPB (Figure 21b). Clearly, the detailed structural comparison of SGPA and alpha-chymotrypsin presented earlier is applicable in most respects to SGPB. However, comment should be made on one aspect of the SGPB structure. The uranyl loop (residues 65A-86) of SGPB is seven residues longer than the same loop in SGPA, but is still much smaller than the uranyl loop of alpha-chymotrypsin. Despite its increased size (Figure 22), the uranyl loop of SGPB takes on a different conformation and occupies a somewhat different surface region than the same polypeptide loop of alpha-chymotrypsin.

D. Active Site Conformation Of SGPB

Figure 23 is a stereo-photograph of several sections of the native electron density map of SGPB (perpendicular to the $\underline{c}$ axis) through the active site region. Residues prominently featured in this map have been labelled according to the numbering scheme of Table 1. Residues in the region of the active site are also illustrated in the

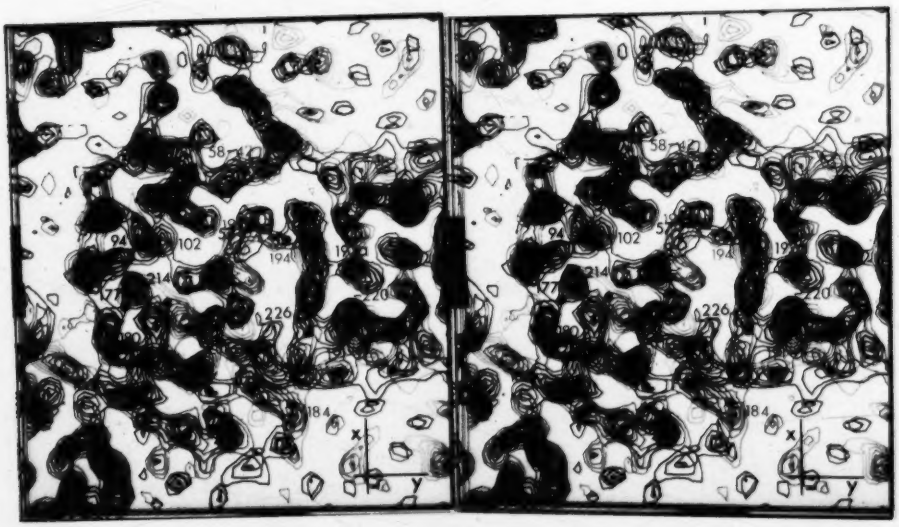


Fig. 23. Stereo-view of eight sections (at one angstrom spacings) of the electron density map of SGPB in the region of the active site. These map sections are perpendicular to the c axis. The first contour is drawn at 0.48e/(angstroms)^3 (including the F(000)/V term of 0.22e/(angstroms)^3) with subsequent intervals increasing by 0.13e/(angstroms)^3. The active site residues Ser-214, Asp-102, His-57 and Ser-195 have been labelled as well as some other residues that are evident on this map. The black dots indicate approximate alpha-carbon atom positions.

stereo-drawing of Figure 24. Reference to Figures 21 and 22 show that the active site residues Ser-214, Asp-102, His-57 and Ser-195 of SGPB have the same geometrical configuration as those of the pancreatic serine proteases and SGPA. Similar to SGPA, there is ordered electron density in close proximity to the side chains of Ser-195 and His-57 in SGPB (Figure 23). These peaks have been tentatively interpreted as bound solvent molecules.

The hydrogen bonding and immediate environment of the catalytic residues in SGPB are also similar to those of SGPA. Asp-102 forms four hydrogen bonds and is in a

hydrophilic environment, albeit this residue is isolated from direct solvent contact by the proximity of a number of residues including His-57 and Phe-94 (Figure 23). The polar environment of Asp-102 in SGPB is a further indication that this residue has a normal pKa rather than an abnormally high pKa as has been suggested (Hunkapiller et al., 1973). His-57, which is hydrogen bonded to Asp-102, appears to form only a weak interaction with the side chain of Ser-195. This is consistent with observations made for other serine proteases (Kraut, 1977). The present interpretation of the 2.8 angstrom map of SGPB indicates a chi1 of -97° for the side chain of Ser-195. This is similar to that found for other serine proteases and SGPA, but is not consistent with the proposed 'up' position (chi1=90°) observed for alpha-chymotrypsin (Blow, 1976).

A number of other active site features are conserved in the structure of SGPB upon comparison with the pancreatic serine proteases and SGPA. For example, the conformation of Ser-214, as discussed earlier, is very similar in all these enzymes (Figure 24). Also similar, is the conformation of the polypeptide chain of residues 193 to 195, which forms the oxyanion hole. Further, as can be seen in Figures 19 and 24, the conformation of the disulfide bridge from residues 42 to 58 is highly conserved in SGPB.

Examination of Figure 22 shows that SGPB does not have a well formed 'tosyl hole' primary binding site as found in alpha-chymotrypsin. In this respect, the primary specificity

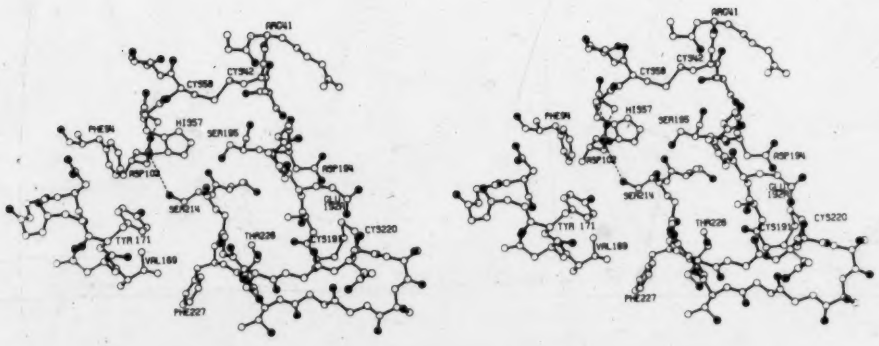


Fig. 24. Stereo-view of the active site region of SGPB looking down the crystallographic c axis. The polypeptide main chain bonding is shown with solid black bonds and oxygen atoms are distinguished by solid black circles. Only hydrogen bonds between active site residues and those with surrounding polypeptide chain have been illustrated as broken lines.

site of SGPB is similar to that of SGPA (see Figure 19). The poorly defined nature of the specificity pocket of SGPB, like that of SGPA, appears to result from a two residue insertion at Ala-192. This insertion results in the rearrangement of the disulfide bridge 191 to 220, causing residues 218 to 220 to turn sharply, allowing only a narrow opening to the primary specificity pocket. This also results in residues 224 to 227 being placed directly under the specificity pocket, making this binding site shallow relative to that of alpha-chymotrypsin. However, one feature of the substrate binding region highly conserved in SGPA, SGPB and alpha-chymotrypsin, is the path of the polypeptide chain of residues 213 to 217 (Figures 21 and 22). This common feature strongly suggests that the microbial enzymes bind substrates in a manner analogous to the pancreatic

enzyme, via an anti-parallel beta structure (Segal et al., 1971; Ruhlmann et al., 1973; Sweet et al., 1974).

Primary specificity site structural differences between SGPB and alpha-chymotrypsin are also reflected in solution substrate studies (Narahashi and Yoda, 1973; Bauer, 1978). These studies indicate the binding pocket of SGPB is less well developed and therefore less specific towards substrate side chains; with side chains as small as alanine and as large as tyrosine being accommodated. As the similarity of their structural features would suggest, this cleavage specificity is also shared by SGPA.

Further solution kinetic studies of SGPB (Gertler, 1974; Bauer, 1978), as found for SGPA (Bauer et al., 1976a,b), indicate enhanced activity occurs with longer inhibitors and substrates. This effect is not as pronounced in alpha-chymotrypsin (Bauer, 1978). As Figures 22 and 24 show, it is likely the proximity of the methionine loop to the expected substrate binding region that is responsible for the formation of additional binding subsites further removed from the active site in the microbial enzymes. In alpha-chymotrypsin, the methionine loop takes on a different conformation out of the active site region (Figure 22). As with SGPA, it is probable that secondary interactions on the surface of SGPB are a necessary prerequisite for proper substrate orientation as a result of less specific binding in the primary specificity site.

V. The Tertiary Structure Of Alpha Lytic Protease At 2.8 Angstrom Resolution

A. Isolation and Crystallization

Alpha lytic protease, generously supplied by Dr. L.B. Smillie, was isolated from culture filtrates of the soil bacillus Myxobacter 495 as described by Whitaker (1965, 1967). The technique of equilibrium dialysis (Zeppezauer et al., 1968) was used to grow suitable single crystals of the enzyme at room temperature (20°C) from 1.3M lithium sulfate at pH 7.2 (the protein concentration was 10 mg/ml). Crystals exhibiting well formed hexagonal prisms with rhombohedral ends of sufficient size (0.4mm x 0.4mm x 0.6mm) for diffractometer study were obtained within a month. A sample of these crystals is shown in Figure 25. An h0l precession photograph of a native enzyme crystal to the limit of 2.4 angstrom resolution is shown in Figure 26. It should be pointed out that crystals of alpha lytic protease grown from lithium sulfate are isomorphous with those grown from 1.7M ammonium sulfate at pH 7.3 (James and Smillie, 1969). The crystal symmetry and unit cell parameters of alpha lytic protease are summarized in Table 14.

The method of Matthews (1968) was used to determine the number of molecules of alpha lytic protease in the asymmetric volume of the crystallographic unit cell. Assuming one enzyme molecule per asymmetric unit and the molecular weight of alpha lytic protease to be 19,869

Fig. 25. Photomicrograph of crystalline alpha lytic protease (40X) grown from 1.3M lithium sulfate at pH 7.2. The long axis of these crystals is parallel to the unique c crystallographic axis.

daltons, as determined by quantitative amino acid analysis (Olson et al., 1970), the value of the volume per unit molecular weight, Vm, can be calculated to be 2.56 (angstroms)^3/dalton. This value of Vm is close to the overall mean value of 2.37 (angstroms)^3/dalton (median value 2.61 (angstroms)^3/dalton) found for a large number of different protein crystals. This result supports the assumption there is only one molecule of alpha lytic protease per asymmetric unit of the crystal lattice. Using these results, it can be further estimated that only 48% of the volume of these crystals contains protein, the remaining volume being made up of solvent channels.

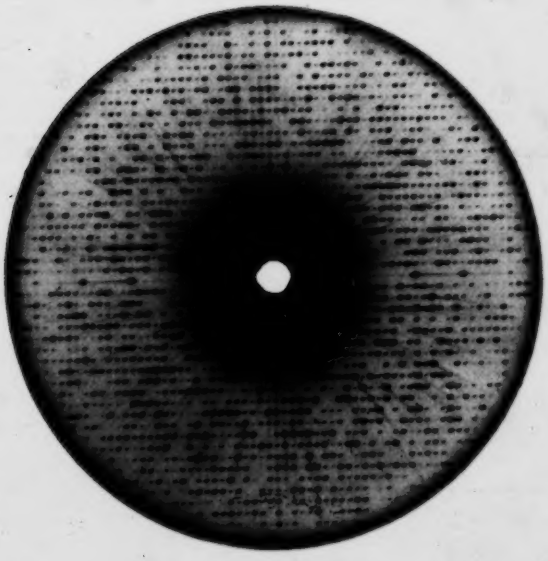


Fig. 26. Precession photograph of a native alpha lytic protease crystal showing the h01 diffraction plane to a limit of 2.4 angstrom resolution. This photograph was taken on a Elliott rotating anode using Ni-filtered Cu K-alpha radiation, 40kV, 20mA and 24 hours exposure time.

B. Data Collection

The equipment and methods used to collect X-ray intensity data from crystals of native alpha lytic protease and its heavy-atom derivatives are described in Table 15. The methodology employed is more fully documented in the SGPA experimental section.

C. Heavy-Atom Derivatives

Five usable isomorphous heavy-atom derivatives of native alpha lytic protease crystals were elucidated in the same manner as described for crystals of SGPA. These derivatives and their optimal soaking conditions are listed

TABLE 14

Crystal Data For Alpha Lytic Protease

Unit cell dimensions	
a	66.32(2) angstroms
b	66.32(2)
c	80.10(2)
V	3.05 x 10^5 $(angstroms)^3$
Unit cell content	6
Systematic absences	001:l=3n±1
Space group	P3(2)21
Growth conditions	1.3M lithium sulfate, pH 7.2

in Table 16. Crystals of alpha lytic protease and its heavy-atom derivatives were sufficiently resistant to radiation damage to allow the collection of a unique hextant of diffraction data to 2.8 angstrom resolution. However, the number of additional Friedel mate reflections collected from a given derivative crystal, was based on the rate of decay and fluctuation in the monitor reflections of that crystal. Unexpectedly, derivative crystals prepared with mercuric chloranilate were remarkably stable to irradiation (Table 16).

Unit cell dimensions for the native and heavy-atom derivative crystals were determined from the centered two-theta positions of six reflections in the range 20° < two-theta < 27° and the positions of their minus two-theta

TABLE 15

Data Collection Methods

Diffractometer: Picker FACS-1, PDP8L computer control.

Operating system: Vandy data collection system written by Lenhert (1975).

Incident beam: Ni-filtered Cu K-alpha; tube operated at 40kV and 24mA.

Diffracted beam: 65cm crystal to counter, beam path was helium filled.

Scan type: omega scan, continuous.

Scan width, speed: 0.5° at 2.0°/min.

Backgrounds: Two 4 second fixed position counts taken 0.4° offset from omega=0° in omega direction.

Net time/reflection: about 40 seconds/reflection

Temperature: ambient 15±2°C

Friedel reflections: measured with counter arm at minus two-theta.

Friedel pair mates. Changes in the unit cell dimensions of derivative crystals from those observed for the native protein crystal are shown in Table 17. The very small changes in unit cell dimensions (not greater than 0.2%) indicate the high degree of isomorphism between native and derivative crystals.

TABLE 16

Statistics of Data Collection to 2.8 Angstrom Resolution For Alpha Lytic Protease

Data Set	Native	Plat[1]	PMA[2]	Plat+ PMA	MC[3]	MC+ PMA
Soaking time/concentration	-	3d/5mM	16h/sat.	3d,16h/ 5mM,sat.	12h/sat.	12h/sat., sat.
Total reflections measured	11208	11234	11298	8618	8259	7231
No. of independent reflections	5305	5235	5267	5320	5113	5017
Maximum absorption correction (%)	10.4	20.8	25.9	12.3	13.4	6.4
Maximum crystal decay (%)	3.4	4.4	11.3	28.2	1.8	1.4
$R(sym)=(\sum_i \lvert\bar{I}-I\rvert_i / \sum_i I_i) \times 100$ (%)	2.7	5.9	5.6	6.0	7.7	5.7
No. of reflections merged	5273	140	140	140	136	131
%Reflections with {I>3(sigma(I))}	94.2	87.6	81.5	88.5	91.8	95.3

[1] Platinum diamino dichloride
[2] Phenylmercuric acetate
[3] Mercuric chloranilate

TABLE 17

Cell Dimension Changes in Derivative Crystals[1]

Data Set	a (sigma a)	%change	c (sigma c)	%change
Native	66.32 (2)	-	80.10 (2)	-
Plat	66.32 (4)	0.00	80.12 (4)	+0.02
PMA	66.41 (4)	+0.14	80.26 (5)	+0.20
Plat + PMA	66.31 (2)	-0.02	80.12 (3)	+0.02
MC	66.30 (3)	-0.03	80.09 (4)	-0.01
MC + PMA	66.31 (2)	-0.02	80.12 (2)	+0.02

[1]a and b were constrained to the same value due to crystal symmetry. All unit cell dimensions are in angstrom units. Sigma values represent the precision of a single determination from one crystal.

D. Data Reduction And Scaling

Results of the complete data reduction from intensities to structure factor amplitudes for native and derivative crystals of alpha lytic protease are given in Table 16. Reflection backgrounds were adjusted as they had been for SGPA. An absorption correction was applied (North *et al.*, 1968) based on a single averaged absorption curve derived from two 001 reflections (l=9,18). The large absorption correction and the fact that relatively few reflections had I > 3sigma(I) for the PMA derivative is probably the result of a smaller and more tabular crystal. Linear decay corrections were determined by monitoring three reflections: 4,0,2; 0,2,20; and 0,0,18 after every 100 reflections collected. Following decay correction, all symmetry

equivalent reflections in the native data set were merged. For heavy-atom derivative data sets, only those symmetry equivalent reflections other than Friedel mates were averaged. In order to derive structure amplitudes, standard Iorentz and polarization corrections were made and the square roots of the resultant intensities calculated.

For all diffraction data sets collected, an absolute scale and overall isotropic thermal B were calculated following procedures used for SGPA. The overall isotropic temperature factor calculated for the native alpha lytic protease crystal was 9.7 (angstroms)^2. This compares very favorably with those for other proteins in a similar molecular weight range. From these calculated Wilson plot absolute scale factors, each heavy-atom derivative data set was scaled to the native enzyme data set as described in the structure solution of SGPA. Reflections with I > 3sigma(I) were excluded from the data sets at this stage. After scaling, heavy-atom differences were then calculated. Table 18 shows the statistics of data scaling and the resultant heavy-atom differences observed.

E. Phase Angle Determination

By way of a three-dimensional heavy-atom difference Patterson map, coordinates for the single major mercury site of the PMA derivative were determined. The subsequent refinement of the positional and thermal parameters of this heavy-atom site provided initial protein phases. Major

TABLE 18

Data Set Scaling And Heavy-Atom Differences

Data Set	Native	Plat	PMA	Plat+ PMA	MC	MC+ PMA
Wilson plot absolute scale	20.62	25.84	29.59	23.92	17.12	18.45
Overall isotropic B (angstroms)^2	9.7	13.4	14.0	17.9	8.9	8.8
Abs. scale (D) / Abs. scale (N)	-	1.253	1.435	1.160	0.830	0.895
Linear scale ratio	-	1.257	1.439	1.172	0.809	0.830
Heavy-atom difference factor R(D)	-	0.123	0.178	0.161	0.218	0.240

heavy-atom sites of the other derivatives were then elucidated by way of the cross Fourier technique (Dickerson et al., 1967). Further heavy-atom site refinement and protein phase determination followed the methodology used in the structural analysis of SGPA. The correct enantiomorphic space group was determined at an intermediate stage in phase determination. The same set of heavy-atom coordinates for three of the five derivatives eventually used, produced a <m> of 0.61 in space group P3(1)21 and a <m> of 0.69 for P3(2)21. On this basis, the space group P3(2)21 was chosen in further computations. This choice of space group was subsequently confirmed by the fit of L-amino acids in the final native electron density map. Several minor heavy-atom sites for the various derivatives were elucidated by the double difference Fourier technique, during the process of heavy-atom site refinement and native protein phase determination. Only data from 10.0 to 2.8 angstrom resolution were used in heavy-atom refinement since lower resolution data agreed poorly with calculated values.

The final heavy-atom site parameters of the five heavy-atom derivatives used in this study are listed in Table 19. In this table sites with negative occupancies indicate the displacement of native electron density upon heavy atom binding. The variation of the ratio r.m.s. f(H), the heavy-atom scattering amplitude, to r.m.s. E(H), the lack of closure error, as a function of $\{\sin(\theta)/\lambda\}^2$ is illustrated in Figure 27. The variation of the average

TABLE 19

Refined Heavy Atom Parameters For Alpha Lytic Protease Derivatives

Derivative	Site	x/a	y/b	z/c	A[1]	B[2]
Plat	1	0.1577	0.3747	0.3051	32.6	9.9
	2	0.1340	0.3255	0.3056	5.9	20.0
	3	0.1844	0.3929	0.2819	10.8	16.2
PMA	1	0.0549	0.4939	0.5409	46.0	13.7
	2	0.1795	0.5054	0.5627	5.2	1.0
	3	0.8958	0.4693	0.0631	7.3	9.8
	4	0.9307	0.4526	0.1566	-7.1	7.6
	5	0.4796	0.5063	0.2235	-5.1	1.0
MC	1	0.3700	0	2/3	72.6	11.4
	2	0.8147	0.3334	0.6965	30.7	13.7
	3	0.7097	0.4236	0.7065	26.2	9.6
	4	0.0698	0.4401	0.3248	9.1	13.0
	5	0.5035	0.2066	0.3585	8.4	10.0
	6	0.5990	0.2653	0.3782	6.6	14.6
	7	0.5377	0.4118	0.7323	6.2	21.0
	8	0.8923	0.4596	0.0839	5.7	10.0
	9	0.1410	0.3913	0.2816	4.8	17.9
	10	0.1346	0.2413	0.0988	-4.4	1.0
	11	0.1070	0.4031	0.2960	-3.1	11.8
	12	0.2280	0.4191	0.1158	-2.8	10.0
PMA + Plat	1	0.1561	0.3736	0.3035	37.8	16.7
	2	0.1367	0.3276	0.3104	12.7	23.0
	3	0.1886	0.3922	0.2765	12.7	10.8
	4	0.8276	0.2350	0.3769	11.6	11.1
	5	0.1666	0.3668	0.2712	11.2	5.6
	6	0.2126	0.4303	0.2792	11.2	10.8
PMA + MC	1	0.3693	0	2/3	69.6	9.2
	2	0.6662	0.4815	0.3630	36.0	9.3
	3	0.7071	0.4222	0.7058	33.4	9.4
	4	0.0564	0.4958	0.5424	14.4	8.7
	5	0.5017	0.2082	0.3559	13.5	8.3
	6	0.8825	0.4583	0.0747	9.5	9.3
	7	0.6040	0.2642	0.3785	7.1	7.9
	8	0.1336	0.2394	0.0955	-4.3	1.0

[1]A is the site occupancy on an approximately absolute scale in electrons.
[2]B is the isotropic temperature factor coefficient, in units of (angstroms)^2.

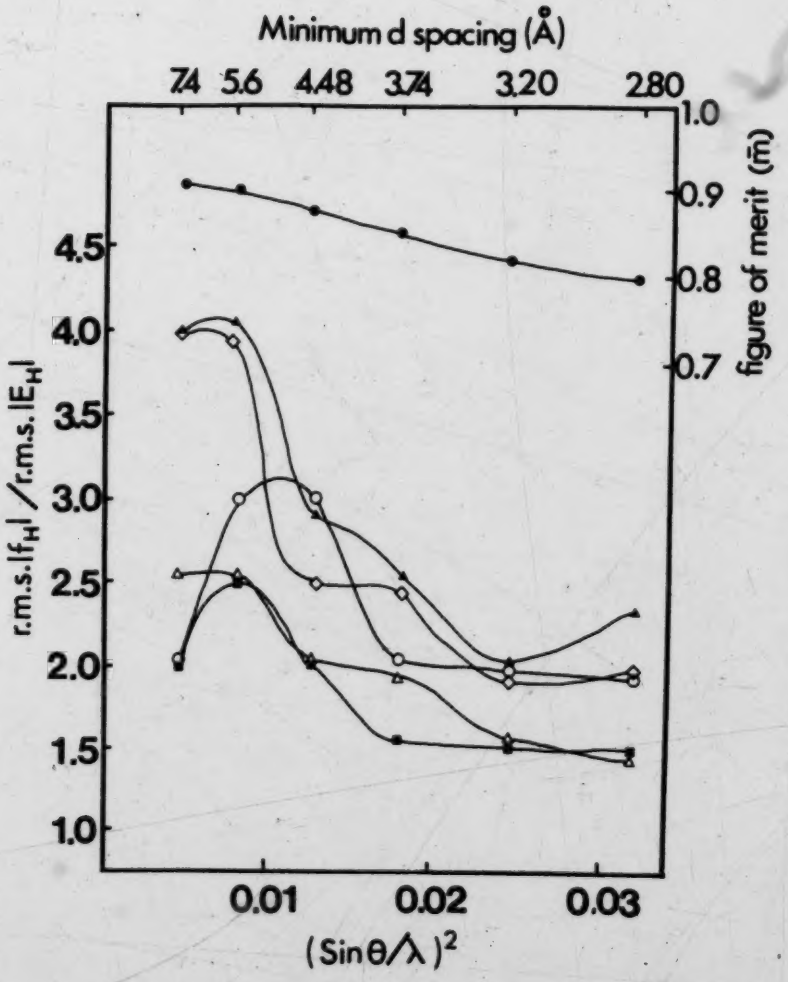


Fig. 27. The variation of the ratio r.m.s. f(H) to r.m.s. E(H) and the mean figure of merit, as functions of {sin(theta)/lambda}^{2}. The derivatives are represented by the following symbols: (o) Plat; (Δ) phenylmercuric acetate; (◇) mercuric chloranilate; (■) phenylmercuric acetate and Plat; (▲) phenylmercuric acetate and mercuric chloranilate. The uppermost curve and the scale to the right show the variation of the figure of merit, using the symbol (●).

figure of merit with resolution is also shown in Figure 27. A histogram of the distribution of figure of merits among the measured native enzyme reflections, is shown in Figure 28. Of the 4866 native reflections for which phases were

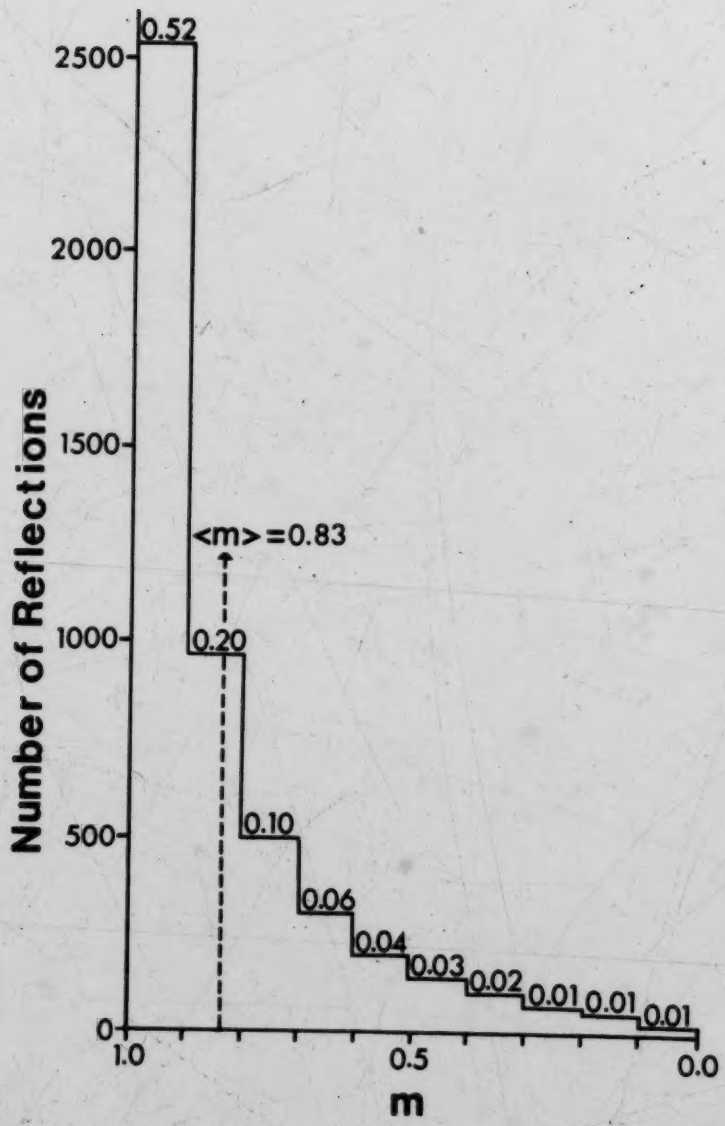


Fig. 28. The distribution of figures of merit among native enzyme reflections of alpha lytic protease. The precentage of the total reflections falling into each range is shown at the top of each column. The dashed line indicates the overall mean figure of merit for all reflections.

TABLE 20

Phase Determination Statistics For Alpha Lytic Protease[1]

Derivative	R(D)[2]	R(c)[3]	R(k)[4]	[5]r.m.s. f(H) /r.m.s. E(H)
Plat	0.123	0.476	0.054	2.251
PMA	0.178	0.580	0.091	1.810
Plat + PMA	0.161	0.586	0.084	1.722
MC	0.218	0.510	0.100	2.360
MC + PMA	0.240	0.491	0.100	2.590

[1]The overall mean figure of merit was 0.83.
[2]R(D) is the heavy-atom difference R-factor.
[3]R(c) is the Cullis R factor (Cullis et al., 1961).
[4]R(k) is the Kraut R-factor (Kraut et al., 1962).
[5]Over all the reflections phased.

determined, more than 92% have an $m \geq 0.5$. The average phase angle difference, computed from the equation $|\text{alpha(max)} - \text{alpha(best)}|$, (Blow and Crick, 1959) for the native reflections phased was 11.7°. Other standard heavy-atom refinement and phase determination statistics are given in Table 20. The overall average figure of merit for the phase determination of the native structure amplitudes of alpha lytic protease was 0.83, indicating a high level of confidence in the structural analysis of this enzyme could be expected.

TABLE 21

Major Sites of Heavy Atom Binding

Heavy-atom site	Nearby protein atoms	Distance[1]
(a) Plat: Site 1	Cys-137 SG	3.4
PMA + Plat: Site 1	Cys-159 SG	1.9
(b) PMA Site 1	His-57 NE2	3.4
	Ser-195 OG	3.6
	Ser-214 O	2.3
(c) MC: Site 1	Thr-143 O	3.8
MC + PMA: Site 1	Gln-158 NE2	3.6
	Gln-158 OE1	3.5
	Cys-220 SG	4.0
	+ the same 2-fold related atoms of another molecule	
(d) MC: Site 2	Arg-103 NEH1	2.5
MC + PMA: Site 2	Pro-233 O	2.7
	Gln-236 NE2	3.9
(e) MC: Site 3	Arg-48B NEH1	2.0
MC + PMA: Site 3	Gln-236 O	3.0
	Tyr-237 O	3.3

[1]All distances are in angstrom units.

F. Heavy-Atom Binding Sites

Heavy-atom - protein interactions of the five heavy-atom derivatives used (Table 19) to solve the alpha lytic protease structure have been studied in detail, using protein atomic coordinates derived from the interpretation of the native enzyme map. Protein atoms near major heavy-atom binding sites, along with the interaction distances involved, are given in Table 21. Similar sites in different heavy-atom derivatives have been grouped together in this Table. The major platinum diamino dichloride binding

TABLE 22

Minor Sites of Heavy Atom Binding

Derivative	Site[1]	Description
Plat	2,3	Probably chlorine atoms bound to a central platinum atom (Site1), from which they are 2.8 and 2.4 angstroms, respectively.
PMA	2	Attachment site near C-terminus.
	3	Small population of dislocated His-57 side chains.
	-4	Displacement of solvent near Ser-195 and major PMA site.
	-5	Centered on native His-57 position indicating a certain population has been dislocated, probably to Site 3.
MC	4	Interacts with the polypeptide backbone of residues 143 and 147.
	5	Near Arg-48B, opposite side of side chain to major site 3.
	6	A further small peak at Arg-48B.
	7	Near the side chain of Glu-174.
	8	Near alternate side chain oxygen of Glu-174 - see site 7.
	9	Close to disulfide bridge 137-159.
	-10	Centered on side chain of Val-132.
	-11	Positioned on main chain Gln-158.
	-12	Near the side chain of Glu-129.
Plat + PMA	2,3	Same as Plat sites 2 and 3
	4,5,6	Sites near disulfide 137-159.
MC + PMA	4	Same as major PMA site 1
	5,6,7,-8	Sites corresponding to the mercuric chloranilate sites 5,8,6,-10, respectively.

[1]A site denoted by a negative number indicates a negative occupancy (displacement of native electron density).

site is within a covalent bond distance of the SG atom of Cys-159. This fact, plus the observation that this is also a relatively noisy region in the double difference map, could indicate heavy-atom binding had disrupted the disulfide

bridge, Cys-137 to Cys-159. However, this has not been experimentally confirmed. Table 22 briefly describes binding site environments found for the minor sites of each derivative.

G. Native Electron Density Map And Interpretation

The final MIR phased native electron density map was computed using 'best' phases and the figures of merit found for the 4866 native reflections to 2.8 angstrom resolution. The electron density function was sampled along a, b and c at intervals of 0.75, 0.75 and 1.00 angstroms, respectively. Sections of electron density were computed perpendicular to the c crystallographic axis, and then plotted on a scale of 3mm/angstrom. This map included the following region of the unit cell: x (0.11, 1.00); y (0.00, 0.99); and z (-0.12, 0.50). The standard error of the native electron density was estimated to be 0.159e/(angstroms)^3 (Cruickshank, 1949; Dickerson et al., 1961). The portion of standard error due to the measurement of structure factors alone was 0.006e/(angstroms)^3.

The first contour was drawn on the native map to represent an electron density of 0.41e/(angstroms)^3 (including 0.21e/(angstroms)^3 contributed by the F(000)/V term). Further contour lines were placed at 0.10e/(angstroms)^3 intervals. Using these contouring conditions, all portions of the polypeptide chain of alpha lytic protease were clearly visualized. The largest peaks of

electron density occurred at sulfur atom positions (the average value being 1.21e/(angstroms)^3). Using known polypeptide sequence information (Olson et al., 1970) and the location of bound heavy-atoms, it was possible to assign initial alpha-carbon positions in appropriate electron density for all 198 amino acid residues of alpha lytic protease.

The detailed atomic interpretation of alpha lytic protease at 2.8 angstrom resolution was made using a larger scale map (2cm/angstrom) in an optical comparator (Richards, 1968). This process was guided by the alpha-carbon coordinates determined from the previous smaller scale map. The optical comparator map was calculated in sections perpendicular to the c crystallographic axis with dimensions: x (0.22, 0.93); y (0.08, 0.93); z (-0.07, 0.40) and grid intervals of 0.75 angstroms in all axial directions. The first contour was drawn to represent an electron density of 0.31e/(angstroms)^3 and subsequent contours were drawn at intervals of 0.10e/(angstroms)^3. Watson-Kendrew skeletal units were then connected to depict the chemical sequence and manipulated into the electron density distribution to achieve the final optimal fit of the atomic model of alpha lytic protease.

Coordinates for all non-hydrogen atoms were measured from this model using the plumb-line method. These coordinates were used as guide points to obtain the best fit of a stereo-chemically correct structure with standard bond

lengths and interbond angles (Diamond, 1966, 1974). The final r.m.s. deviation of the idealized structure from the original measured coordinates was 0.21 angstroms for all 1391 non-hydrogen atoms in the molecule.[7]

H. Atomic Model Of Alpha Lytic Protease

An alpha-carbon atom stereo-drawing of alpha lytic protease as interpreted from the 2.8 angstrom resolution native electron density map is shown in Figure 29a. Each alpha carbon atom position in Figure 29a has been designated according to the numbering scheme of Table 1. The polypeptide chain of alpha lytic protease is folded to form two structurally distinct hydrophobic domains, with their juncture forming a surface depression in which the active site is located. The apparent deep pocket, bound by five stretches of polypeptide chain (residues 190-194, 137-140, 197-199, 213-217 and 226-228), is in reality filled with amino acid side chains which makes this a relatively small active site surface depression. Each domain (one composed of the N-terminal portion of the polypeptide chain; the other of the C-terminal portion) is composed of four anti-parallel beta loops folded to form a beta barrel type structure (Birktoft and Blow, 1972). Those loops forming the N-terminal domain are the N-terminal loop (residues 15A-41), the histidine loop (residues 42-58), the uranyl loop

---

[7]Final non-hydrogen atom coordinates for SGPA, SGPB and alpha lytic protease have been deposited in the Protein Data Bank at the Brookhaven National Laboratory, U.S.A.

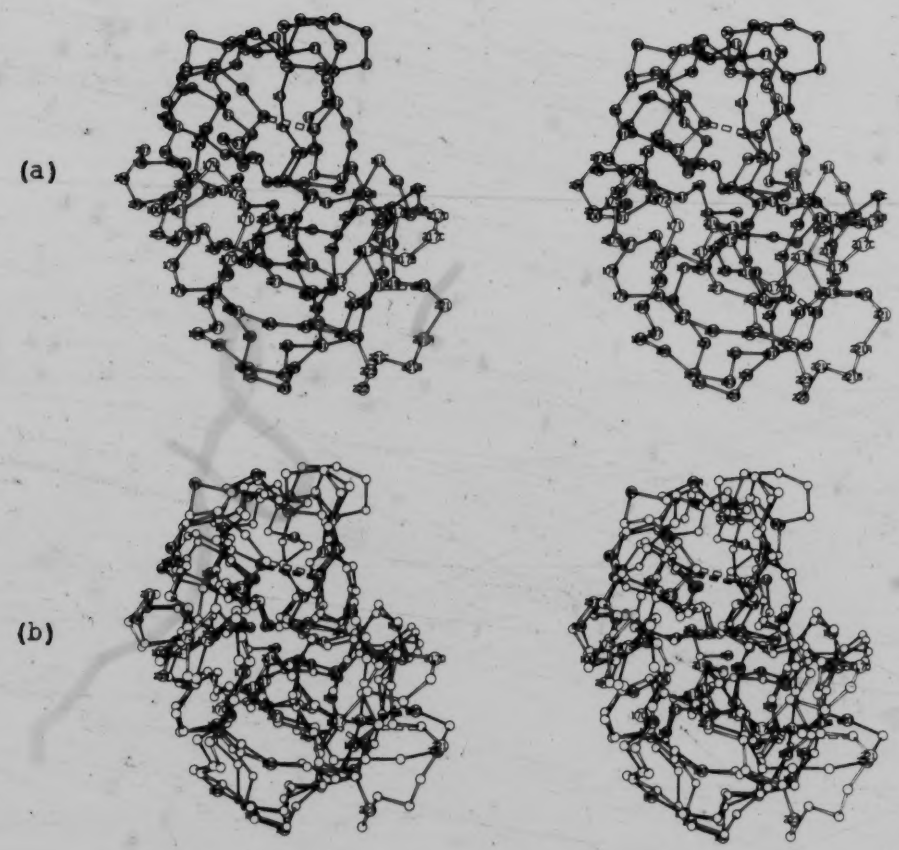


Fig. 29:
(a) A stereo-drawing of the alpha-carbon backbone of alpha lytic protease, with each alpha-carbon position numbered according to Table 1. The active site is located in the central portion of this drawing, where alpha-carbon positions of the four active site residues: Ser-214, Asp-102, His-57 and Ser-195 are evident. The three disulfide bridges 42-58, 137-159 and 191-220 present in this molecule are denoted by dashed virtual bonds.
(b) The alpha-carbon drawing of alpha lytic protease is superimposed on that of SGPA in a manner designed to maximize topological equivalence. Regions of similar tertiary structure between these two enzymes are clearly evident, as are areas of structural differences. Every fifth alpha-carbon position of each enzyme is numbered; additional numbering is also present for residues specifically discussed in the text. The two disulfide bridges (42 to 58 and 191-220) of SGPA are denoted by dashed black virtual bonds.

(residues 65-86) and the aspartate loop (residues 87-107). Loops of the C-terminal domain are the autolysis loop (residues 132-163), the methionine loop (residues 164-182), the serine loop (residues 195-213) and the specificity loop (residues 214-228). This overall double beta barrel structure is a common feature of SGPA, SGPB and the pancreatic serine proteases (Birktoft and Blow, 1972; Sawyer et al., 1978). The polypeptide chain folding of alpha lytic protease and all of these other serine proteases can be represented in a fashion similar to that depicted in Figure 20.

The polypeptide backbone conformation of alpha lytic protease is represented in the phi, psi plot of Figure 30 (Ramakrishnan and Ramachandran, 1965). Most of the amino acid residues have phi, psi angles in the region corresponding to a beta pleated sheet conformation. Indeed, Figure 29a shows that there is only one short helical segment of polypeptide chain (about 2 turns) in alpha lytic protease (residues 231-238). An earlier study (Paterson and Whitaker, 1969) also demonstrated the lack of helical structure in this enzyme. At the present resolution the nature of this helical segment cannot be distinguished definitively between 3(10) or alpha helix.

The peptide bond joining Phe-94 and Pro-99A of alpha lytic protease was found to be a cis-peptide link. Cis-Pro-99A is located at a hairpin bend (see Figure 31) and assumes a conformation very similar to that found for this

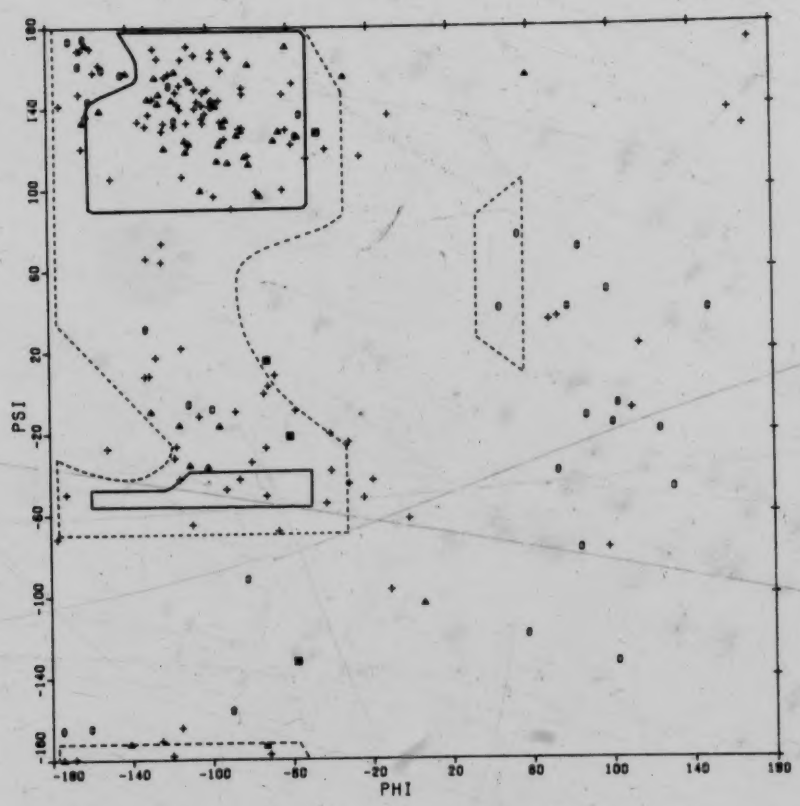


Fig. 30. Plot of the phi, psi torsional angles for the atomic model of alpha lytic protease. The area enclosed within the solid lines of this plot is the fully allowed conformational region for tau{C(alpha)} of 110°, whereas the broken line indicates the outer limit of acceptable van der Waals' contacts for tau{C(alpha)} of 115°. The symbols used represent the following amino acids: (▲) beta branched amino acids; (o) glycine; (■) proline; (+) other amino acid residues.

same residue in SGPA (Figure 10) and SGPB. This is a common feature found at the hairpin bend of the aspartate loop (residues 87-107), in all pancreatic-like microbial serine proteases for which tertiary structures have been solved.

Secondary structural features of alpha lytic protease are shown in the main chain hydrogen bonding diagram of

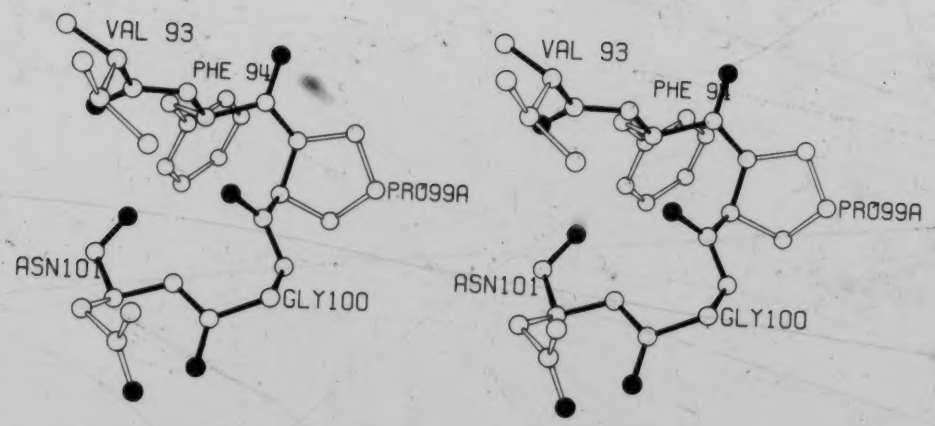


Fig. 31. A stereo-drawing of the Phe-94 to Pro-99A cis-peptide link at the beta bend extremity of the aspartate loop in alpha lytic protease. Comparison of this drawing with that of Figure 10 shows the similarity of this feature to the cis-peptide bond found in SGPA.

Figure 32. In a stylized fashion, this figure like that for SGPA (Figure 11), shows the general polypeptide folding of alpha lytic protease. Figure 32 also depicts which portions cf polypeptide chain are in sufficiently close proximity to form the anti-parallel beta sheet structures of each hydrophobic core of the enzyme. The majority of the main chain hydrogen bonds illustrated are intra-domain, with only a few of them linking the two hydrophobic cores together. Close contacts were designated as hydrogen bonds based on the criteria set earlier for hydrogen bonds in SGPA. Comparison of Figure 32 with a similar diagrammatic representation of hydrogen bonding for SGPA (Figure 11), shows the remarkable conservation of polypeptide chain folding and hydrogen bonding in these two enzyme structures.

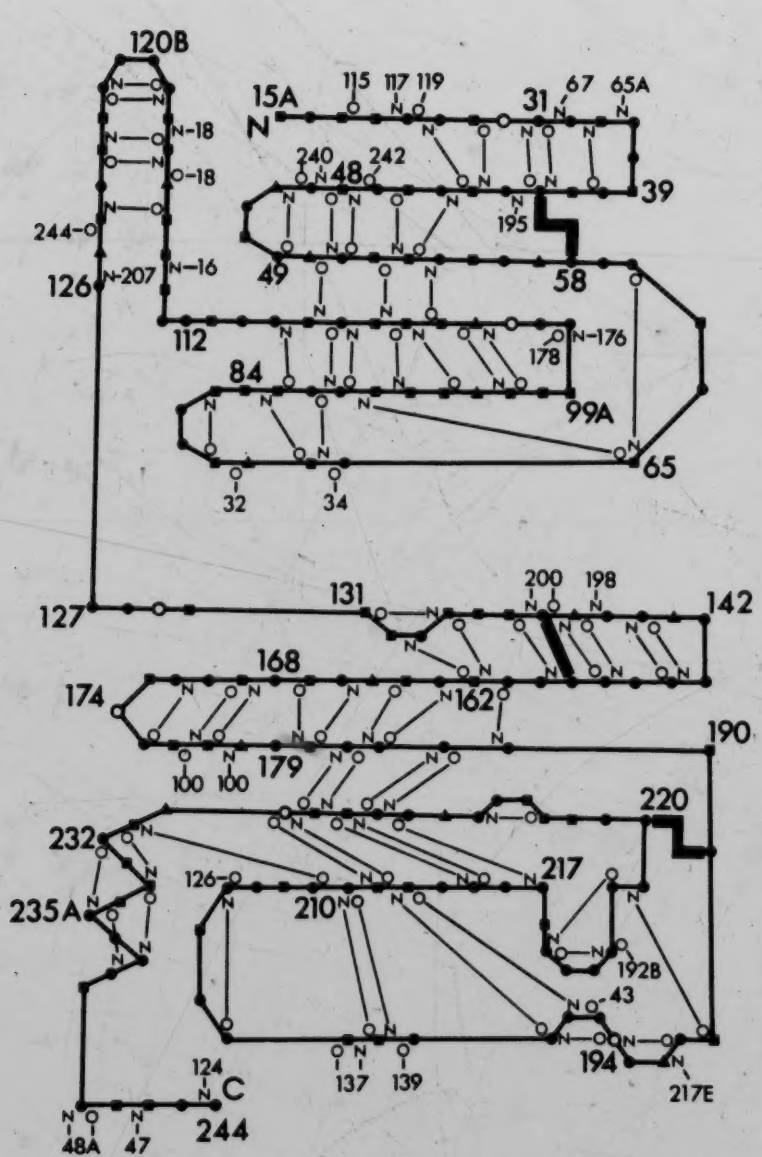


Fig. 32. Schematic drawing of the observed secondary structural features of alpha lytic protease. All observed hydrogen bonds between main chain carbonyl oxygen and imino nitrogen atoms are indicated. Charged acidic residues are denoted by (o), basic groups by (▲), hydrophilic uncharged by (●), and hydrophobic by (■). The three disulfide bridges present are shown as thick black lines.

Further comparisons of the hydrogen bonding schemes of the pancreatic enzymes alpha-chymotrypsin (see Figure 7 of Birktoft and Blow, 1972) and elastase (see Figure 9 of Sawyer et al, 1978), with those of the microbial serine proteases, reveals the conservation of polypeptide chain folding patterns in these two protease groups despite their low sequence homology. Also, a number of highly conserved hydrogen bonds can be observed in all of these enzymes, particularly near functionally important amino acid residues.

Other hydrogen bonds which stabilize the tertiary structure of alpha lytic protease are listed in Table 23 (main chain - side chain) and Table 24 (side chain - side chain). Salt bridges between oppositely charged side chains in Table 24 have been indicated by an asterix. The side chains of Arg-138 and Asp-194 are in close proximity to each other, but as Table 24 indicates (see Figure 34), well formed hydrogen bonds are not formed between these two residues.

Residues of the polypeptide chain involved in the formation of beta bends of types I(10) and II(10) (Venkatachalam, 1968) are summarized in Table 25. Other less well defined beta bends which do not fulfill these requirements, involve residues 33 to 40, 48B to 49, 94 to 100, 140 to 156 and 171 to 175.

TABLE 23

Main Chain to Side Chain Hydrogen Bonds

Ala-15A	N	Gln-112	OE1	Leu-115	O	Ser-47	OG
Ser-43	O	Thr-54	OG1	Ala-119	O	Ser-139	OG
Gly-45	O	Ser-198	OG	Ala-130	N	Gln-210	OE1
Arg-48B	O	Thr-49	OG1	Ala-130	O	Lys-165	NZ
Gly-56	N	Asp-102	OD2	Gly-160	O	Asn-184	ND2
Gly-56	O	Arg-91	NE	Ala-190	O	Ser-226	OG
His-57	N	Asp-102	OD2	Met-192	N	Asn-219	OD1
Asn-64	O	Thr-87	OG1	Gly-192A	N	Asp-194	OD1
Thr-65A	N	Asn-34	OD1	Asn-217	O	Asn-219	ND2
Ser-110	O	Lys-50	NZ	Ile-221A	O	Arg-224	NE
Ala-111	N	Thr-109	OG1	Ser-225	N	Asn-217	OD1
Gln-112	O	Lys-50	NZ	Tyr-237	O	Arg-48B	NEH2

TABLE 24

Side Chain to Side Chain Hydrogen Bonds And Salt Bridges

Glu-30	OE1	Arg-141	NEH1*	Arg-125	NE	Gly-244	OT2*
Tyr-31	OEH	Thr-54	OG1	Glu-129	OE2	Arg-230	NEH1*
His-57	ND1	Asp-102	CD1*	Arg-138		Asp-194	*
Asp-102	OD1	Ser-214	OG	Thr-142	OG1	Asp-194	OD2
Arg-103	NE	Glu-229	OE1*	Thr-143	OG1	Gln-158	OE1
Arg-125	NEH2	Gly-244	OT1*	Tyr-171	OEH	Ser-214	OG
Asn-217	ND2	Gln-217B	OE1				

I. Structural Comparison Of Alpha Lytic Protease With SGPA And Other Serine Proteases

The initial discovery of significant primary sequence homology in active site sequences of alpha lytic protease and pancreatic elastase lead to the proposal that these enzymes would have very similar tertiary structures (Whitaker et al., 1966; Whitaker and Roy, 1967; Smillie and Whitaker, 1967). Elastase is the pancreatic serine protease

TABLE 25

Hairpin Loops Found in Alpha Lytic Protease

Residues	Positions 1	2	3	4	Type
62 - 65	Thr	Val	Asn	Ala	I
80 - 83	Ile	Gly	Gly	Ala	II
120 - 120D	Asn	Gly	Ser	Ser	I
131 - 134	Ala	Val	Gly	Ala	II
192A - 194	Gly	Arg	Gly	Asp	II
194 - 197	Asp	Ser	Gly	Gly	II
201 - 207	Thr	Ser	Ala	Gly	I
217B - 217E	Gln	Ser	Asn	Gly	I
221B - 223	Pro	Ala	Ser	Gln	I

sharing the most similar primary specificity with alpha lytic protease. However, once the full polypeptide sequence of alpha lytic protease became known (Olson et al., 1970) it was found to have relatively little overall sequence homology with elastase. Indeed, as discussed earlier, this lack of sequence homology was primarily responsible for the initial misalignment of some segments of the sequences of these two enzymes. Even with the proper alignment of the sequences of alpha lytic protease and elastase, based upon their known tertiary structures (Table 1), there is only a minimal 18% primary sequence homology. Nevertheless, 55% of the residues of the alpha lytic protease molecule have a topologically equivalent residue in pancreatic elastase (within 2.08 angstroms, see Table 2).

A similar situation is revealed in the comparison of the primary sequence of SGPA and its pancreatic equivalent alpha-chymotrypsin, with which SGPA has only 21% primary

sequence homology (Table 1), but 64% topologically equivalent residues (Table 2). A further comparison of SGPB with alpha-chymotrypsin reveals only 18% primary sequence homology, but 63% topological equivalence. Thus, the class of bacterial pancreatic-like serine proteases, exhibit low primary sequence homology with their pancreatic counterparts, although a much higher level of structural similarity is maintained.

Alpha lytic protease, when compared to SGPA and SGPB, is much more similar with respect to primary sequence homology and tertiary structure, than it is to the pancreatic serine proteases (Table 2). It seems likely that these bacterial enzymes evolved from a common ancestral gene, diverging with respect to primary specificity, to suit the particular requirements of the parent organism from which it was isolated.

A stereo-drawing of the alpha-carbon skeleton of alpha lytic protease and this same drawing superimposed on the structure of SGPA is presented in Figure 29. As Figure 29b shows, there is striking tertiary structural homology between alpha lytic protease and SGPA. A total of 148 residues (82%, Table 2) of SGPA are topologically equivalent to residues in alpha lytic protease within an r.m.s. deviation of 1.46 angstroms. Figure 33 shows two further superimposed stereo-drawings: one of alpha lytic protease and SGPB; the other of alpha lytic protease and its pancreatic counterpart, elastase. The following discussion

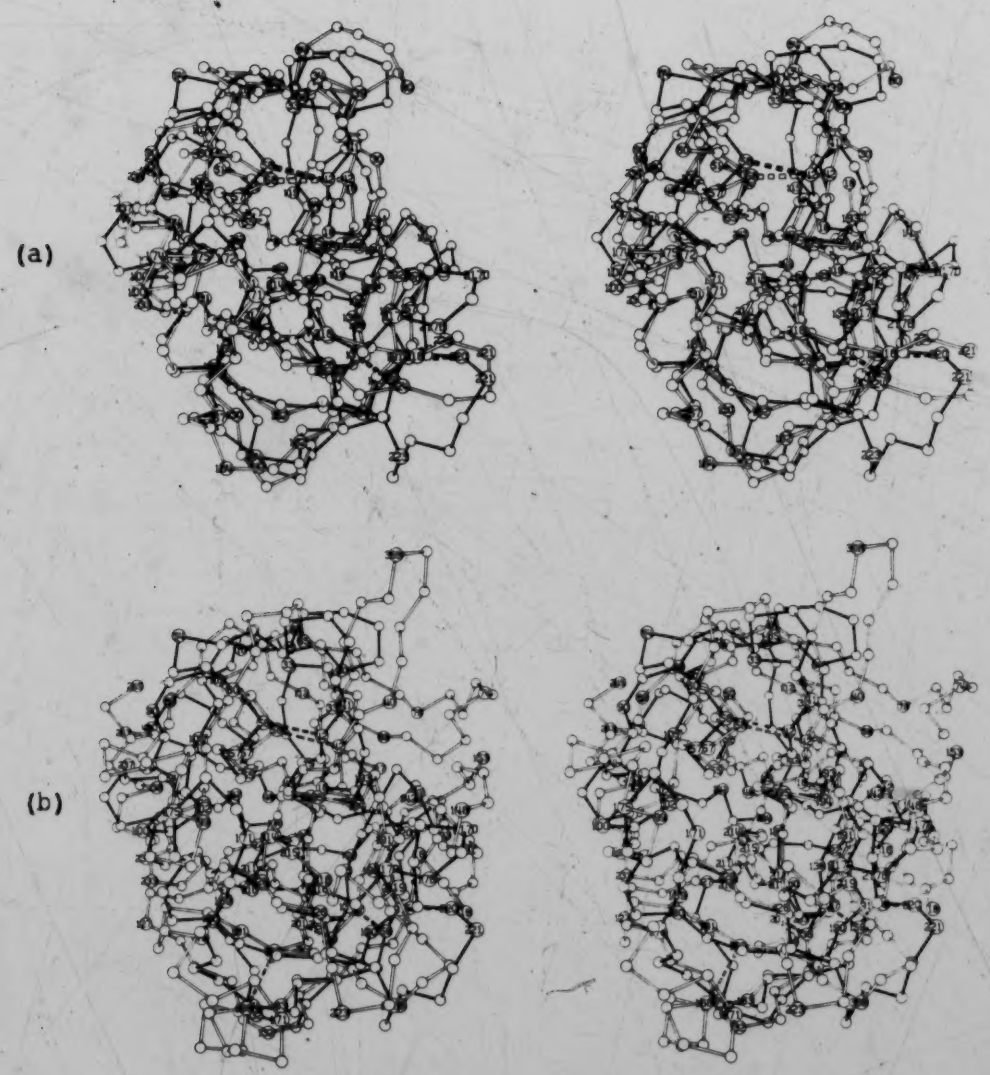


Fig. 33:
(a) A stereo-drawing of the alpha-carbon backbone of alpha lytic protease (black virtual bonds) superimposed on the alpha-carbon backbone of SGPB (open virtual bonds). A total of 154 residues of alpha lytic protease are topologically equivalent to residues of SGPB within an r.m.s deviation of 1.76 angstroms.

(b) In this stereo-drawing the alpha-carbon backbone of alpha lytic protease (black virtual bonds) has been superimposed on the alpha-carbon backbone of elastase (open virtual bonds). Coordinates for elastase were taken from Sawyer et al. (1978). There are 108 residues of alpha lytic protease topologically equivalent to residues of elastase within an r.m.s. deviation of 2.02 angstroms.

is largely limited to a structural comparison of alpha lytic protease and SGPA, since the structures of SGPA and SGPB are so similar. Also, the structural relationship of alpha lytic protease to elastase is very similar to that covered in the previous comparison of SGPA and alpha-chymotrypsin. Nevertheless, specific points of interest pertaining to the comparison of the structure of alpha lytic protease to that of SGPB or of elastase can be observed in Figures 33a and 33b, respectively.

The structure of alpha lytic protease demonstrates that this enzyme, like SGPA, is unlikely to have a zymogen precursor such as those of the pancreatic serine proteases. As shown in Figure 29a the N-terminus of alpha lytic protease does not form a salt bridge to Asp-194 of the active site, the formation of which leads to the activation of the pancreatic enzymes upon zymogen cleavage. Instead, in alpha lytic protease this crucial salt bridge is completed via the side chain of Arg-138. The Arg-138 to Asp-194 salt bridge is buried internally, as in SGPA and SGPB, suggesting its formation occurs upon the initial folding of the polypeptide chain as the enzyme is synthesized. The close similarity of this feature in SGPA and alpha lytic protease can be seen by comparing Figures 16a and 34. Gly-140, also shown in Figure 34, is conserved in all the serine proteases represented in Table 1. Clearly, a residue other than glycine at this position would disrupt the Asp-194 to Arg-138 salt bridge. The conservation of this salt bridge to

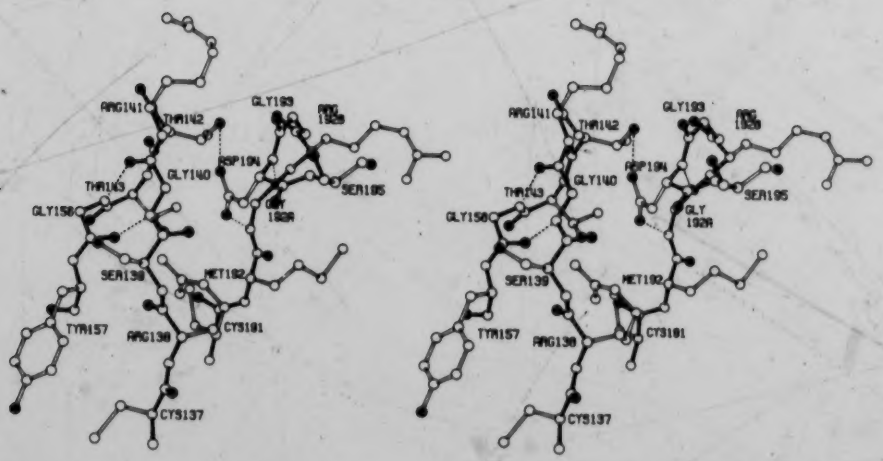


Fig. 34. Stereo-drawing of the environment about the internal salt bridge formed by the guanidinium group of Arg-138 to the active site residue Asp-194 in alpha lytic protease. This structural feature is similar in all microbial pancreatic-like serine proteases for which structures have been elucidated.

Asp-194 in the microbial serine proteases, albeit via a different residue than that of the pancreatic serine proteases, serves to demonstrate the importance that this feature has in determining the enzymatically active conformation of the active site of these enzymes.

The N-terminus of alpha lytic protease is found completely removed from the active site region (the distance from the N-terminus to the carboxyl group of Asp-194 is approximately 21 angstroms). In comparison with SGPA, alpha lytic protease has a two residue insertion in the polypeptide chain at the N-terminus. These residues simply extend the N-terminal polypeptide chain along the surface of the enzyme (Figure 29b) from the N-terminal position observed for SGPA. Figure 35 shows the environment about the

N-terminus of alpha lytic protease and interactions formed with nearby polypeptide chains. Several solvent molecules, although not drawn in Figure 35, are found near the free amino group of Ala-15A. The remainder of the N-terminal loop in both alpha lytic protease and SGPA have a very similar conformation (Figure 29b). A single residue insertion at residue 35 in alpha lytic protease occurs at a surface turn causing no significant polypeptide chain rearrangement.

The histidine loop (residues 42-58), containing the highly conserved disulfide bridge between residues 42 and 58, and the active site residue His-57, has an almost identical conformation in both enzymes. This is also a region of high sequence homology between alpha lytic protease and SGPA, particularly around His-57. Following the histidine loop, alpha lytic protease has a five residue insertion at Ala-66, when compared to SGPA. This insertion occurs at the beta bend of the uranyl loop of SGPA (residues 65A-86), which in that enzyme is very small. The extra five residues found in alpha lytic protease (65 to 83) simply extend this loop along the surface of this enzyme. Thus, the size of the uranyl loop of alpha lytic protease is more like that of SGPB (Figure 33a). However, like SGPB, the uranyl loop of alpha lytic protease is considerably smaller and takes on somewhat different conformation than found in alpha-chymotrypsin (Figure 22) or elastase (Figure 33b).

The aspartate loop (residues 87-107), containing the catalytically important residue Asp-102, is highly conserved

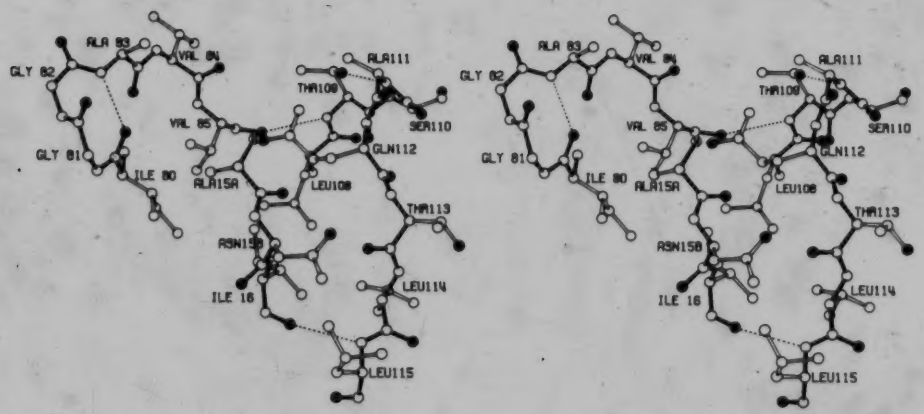


Fig. 35. Stereo-drawing of the environment about the N-terminus of alpha lytic protease showing that it is accessible to acetylation without affecting the conformation of residues in the active site. Several solvent molecules, not drawn here, are found near the N-terminus. Note the beta turn consisting of residues 80 to 83 (type II(10), Table 25).

in tertiary structure in both SGPA and alpha lytic protease. The beta bend of this loop is completed in each enzyme by cis-Pro-99A. As in SGPA and SGPB, the side chain of Asp-102 in alpha lytic protease is isolated from direct solvent contact by the side chains of His-57, Phe-94 and portions of the main chain of the methionine loop. Emerging from the aspartate loop, the polypeptide chain of alpha lytic protease proceeds along the surface opposite to the active site (Figure 29a) and at one point forms a small beta loop (residues 117-124). Alpha lytic protease has one less residue in this minor loop than the equivalent loop of SGPA (Figure 29b), but four more residues than SGPB(Figure 33b).

The first major beta loop of the C-terminal hydrophobic core of alpha lytic protease is the autolysis loop (residues 132-163). It is this loop which contains Arg-138 whose side chain forms the crucial salt bridge to Asp-194 of the active site. As can be seen in Figure 29b, the conformation of this loop is highly conserved in both SGPA and alpha lytic protease. Unlike SGPA and other related microbial and pancreatic serine proteases, alpha lytic protease has a disulfide bridge (137 to 159) which links the two ends of the autolysis loop. If the two topologically equivalent residues of SGPA, Gln-137 and Ser-159, were cysteine residues, they would be close enough to form a disulfide bridge. Overall alpha lytic protease, like SGPA and SGPB, has a much smaller autolysis loop than found in the pancreatic serine proteases.

The methionine loops (residues 164-182) of alpha lytic protease and SGPA, differing in size by only a single residue (Table 1), have very different conformations from that observed for elastase (Figure 33b) and alpha-chymotrypsin (Figure 15). In the microbial enzymes the extended nature of this loop compensates for residues deleted from the aspartate loop, that are present in the pancreatic enzymes. The methionine loops of both microbial enzymes are situated close to possible secondary substrate binding sites similar to those deduced for the pancreatic serine proteases (Segal et al., 1971; Ruhlmann et al., 1973; Sweet et al., 1974). Thus, this loop may function in forming

suitable secondary binding subsites which are more important to substrate binding and in the catalytic hydrolysis of peptide bonds for these microbial enzymes, than it is for their pancreatic counterparts (Bauer et al., 1976b; Bauer, 1978).

The strand of polypeptide chain (residues 183-197) immediately following the methionine loop, contains the highly conserved disulfide bridge at residue 191 and the active site sequence Gly-Asp-Ser-Gly-Gly (residues 193-197). This active site sequence is conserved in all pancreatic-like serine proteases for which sequences have been determined (Table 1). As would be expected, the conformation of these residues in alpha lytic protease is very similar to that observed in SGPA (Figure 29b). The serine loop (residues 195-213), despite the single insertion of Ala-203 in alpha lytic protease, also has a similar conformation in both enzymes.

An interesting structural feature of pancreatic-like serine proteases is preserved in alpha lytic protease. As shown in Figure 36, there are two beta bends about the catalytic residue Ser-195 (Ruhlmann et al., 1973; Birktoft and Blow, 1972; Sawyer et al., 1978). These beta bends are anchored by the disulfide bridge 191 to 220 and the salt bridge formed by Asp-194 to Arg-138. Ser-195 occupies position 2 of the second beta bend (Table 25) from which its side chain is projected into the active site near the other catalytic residues. The highly conserved primary and

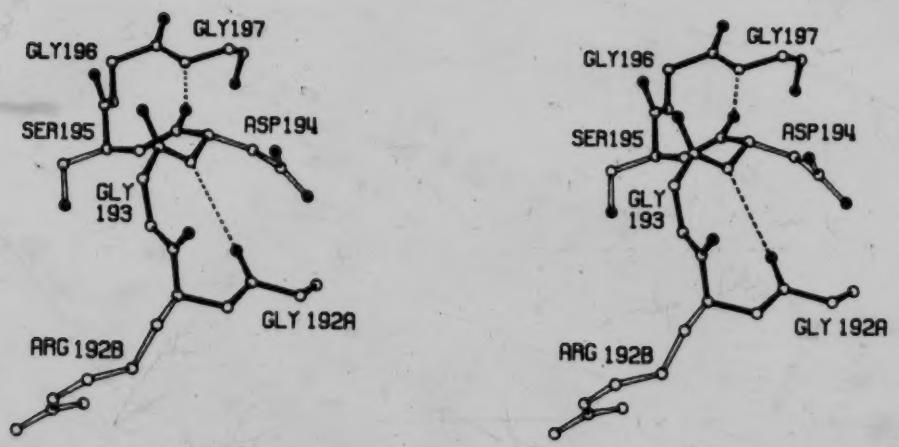


Fig. 36. The tertiary structure of the double beta bend about Ser-195, involving residues Gly-192A to Gly-197, is shown. This structural feature is highly conserved in both pancreatic and pancreatic-like microbial serine proteases. The two beta bend hydrogen bonds are indicated by thin dashed lines.

tertiary structure about Ser-195, indicates the important role this double beta bend structure apparently plays in correctly positioning this active site residue. In the inactive zymogen structures of trypsinogen and chymotrypsinogen A, this stretch of polypeptide chain has a different conformation (Fehlhammer et al., 1977; Freer et al., 1970; Birktoft et al., 1976).

The primary specificity of alpha lytic protease for small amino acid side chains (Ala, Val; Whitaker et al., 1965b) is quite different from that of SGPA, which is specific for aromatic and larger hydrophobic side chains (Phe, Tyr, Leu; Bauer et al., 1976a). As shown in Figure 29b, a five residue insertion into the sequence of alpha

lytic protease at Asn-217 can account, in part, for the primary specificity of this enzyme. These additional residues (Val-217A to Gly-217E) significantly reduce the size of the primary specificity pocket of alpha lytic protease from that observed for SGPA. Despite this insertion, residues Ser 214 to Gly-216 retain a highly homologous structure in both enzymes.

Surprisingly, the five residue insertion at Asn-217 in alpha lytic protease, does not have a dramatic effect on the position of the disulfide bridge, Cys-191 to Cys-220 relative to that in SGPA (Figure 29b). There is a further three residue insertion in the sequence of alpha lytic protease at Gly-221 (Table 1). This insertion is too far removed from the specificity pocket to influence side chain binding. Nevertheless, it does appear to provide the necessary structure to re-establish a more similar conformation to that found in SGPA for the remainder of the C-terminal polypeptide chain of alpha lytic protease. In fact, as shown in Figure 29b, a return to nearly homologous conformations of the polypeptide chain of both enzymes is evident beyond Arg-224.

Following the specificity loop, the remainder of the C-terminal polypeptide chain of alpha lytic protease (residues 229-244) forms approximately two helical turns (Figure 29a), ending with the last six residues in an extended conformation. SGPA forms a similar helical region, but is two residues shorter at the C-terminus. The two extra

residues of alpha lytic protease extend beyond and along the surface from the C-terminal point observed in SGPA.

J. Active Site Conformation

Several sections of the native electron density map of alpha lytic protease, through the region of the active site, are shown in Figure 37. Four residues in the active site region of alpha lytic protease: Ser-195, His-57, Asp-102 and Ser-214, are located in the central portion of this map. These residues retain an overall disposition similar to that observed in all other pancreatic-like serine proteases. Also shown in Figure 37, is a portion of the phenyl ring of Phe-94, which isolates the hydrogen bond interaction between His-57 and Asp-102 from direct solvent contact.

In the active site of alpha lytic protease there is a large electron dense peak near Ser-195, His-57 and the oxyanion hole (residues 193-195). This peak has been tentatively assigned as a well occupied sulfate anion (maximum peak height approximately 1.6e/(angstroms)^3). Only a portion of this peak is visible in Figure 37. It has been possible to position a sulfate anion into this peak, and the interpretation of this sulfate density, drawn with neighbouring amino acid residues of the active site, is shown in Figure 38. The presence of well ordered solvent has been observed in the active sites of several other serine proteases (Kraut, 1977; Sawyer *et al.*, 1978). Such solvent peaks are also found in the active sites of SGPA and SGPB,

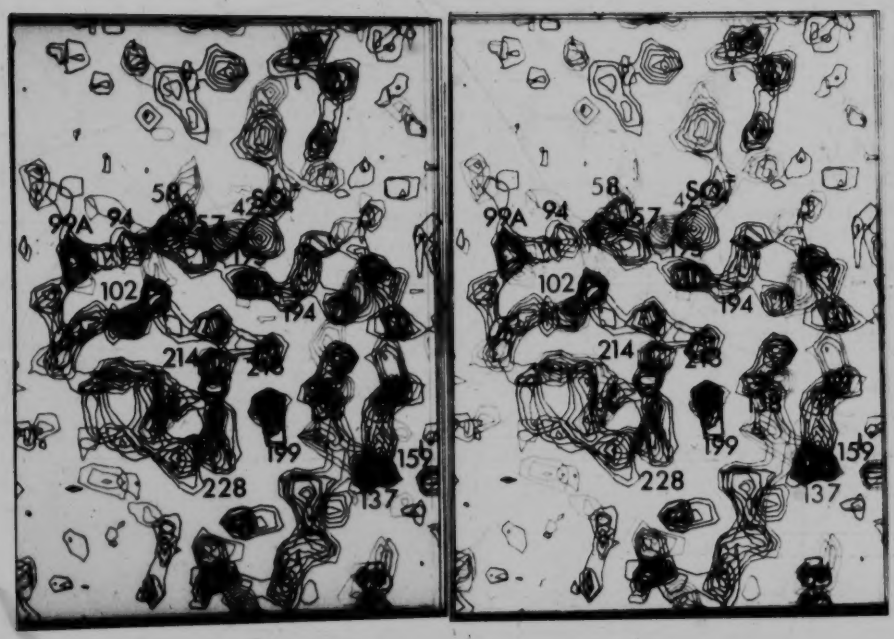


Fig. 37. A stereo-representation of the MIR phased 2.8 angstrom resolution native electron density map of alpha lytic protease through the active site region of the molecule. The first contour is drawn at 0.51e/(angstroms)^3 (including 0.21e/(angstroms)^3 contributed by the F(000)/V term) and subsequent contours are drawn at progressive intervals of 0.15e/(angstroms)^3. The view presented in this map includes a cross section of one complete molecule of alpha lytic protease. Situated in the central portion of the map are four active site residues: His-57, Asp-102, Ser-195 and Ser-214. Also labelled are prominent features which are evident on these sections of the electron density map.

but are much less ordered than the corresponding peak found for alpha lytic protease.

Another feature that can be readily discerned in Figure 37, is the placement of the side chains of Arg-138 and Asp-194. Clearly, this salt bridge in alpha lytic protease is completely internal. Also shown in Figure 37, is the disulfide bridge formed between residues 42 and 58, which is present in all pancreatic-like serine proteases. This bridge

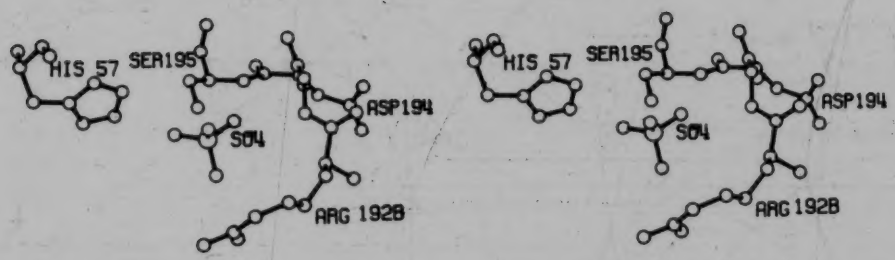


Fig. 38. A stereo-drawing showing the bound conformation of a sulfate anion in the active site of alpha lytic protease. A portion of the electron density peak representing the position of this bound anion is shown in Figure 37.

likely plays an integral role in correctly positioning His-57 in the active site. Towards the lower right side of this map, the disulfide bridge between residues 137 and 159 can be seen. This disulfide bridge is unique to alpha lytic protease and has not been observed in any other pancreatic-like serine proteases. Cis-Pro-99A of alpha lytic protease is partially visible in Figure 37, near the side chain of Phe-94. Several additional features have been labelled in Figure 37; these include Phe-228, Trp-199 and Met-213.

A stereo-drawing of the 2.8 angstrom resolution interpretation of the active site region of alpha lytic protease is shown in Figure 39. This Figure is drawn from a similar vantage point as was used for SGPA, SGPB, and alpha-chymotrypsin (see Figures 19a, 19b, and 24). The conformation of the active site residues Ser-195, His-57,

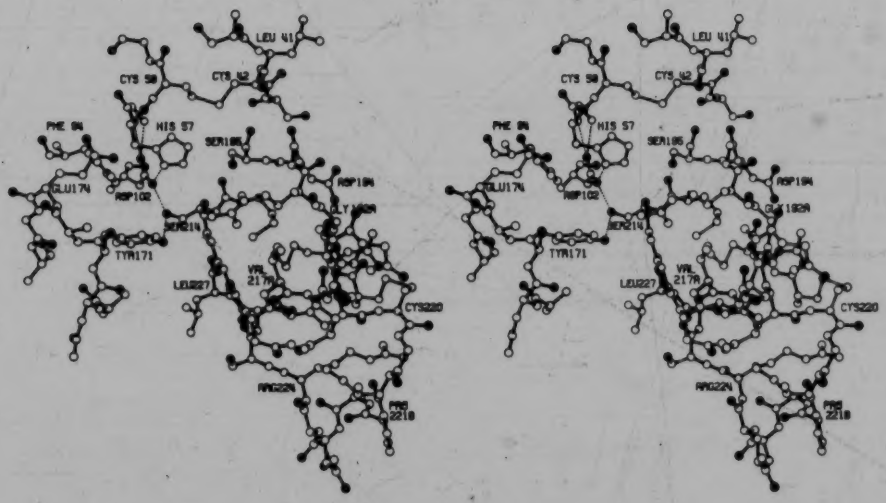


Fig. 39. Stereo-drawing of the active site of alpha lytic protease. The polypeptide main chain bonding is shown with filled in black bonds. All oxygen atoms present are distinguished by filled black circles. Only hydrogen bonds between active site residues and those with surrounding polypeptide chains have been illustrated as dashed lines.

Asp-102 and Ser-214 are very similar in these enzymes, as is the highly conserved disulfide bridge from residue 42 to 58 and the placement of the side chain of Asp-194.

In alpha lytic protease, like other serine proteases, Ser-214 is hydrogen bonded to Asp-102, reaffirming the important role that this highly conserved residue (Table 1) must play in the active site. It is of special interest that this hydrogen bond is to the same Asp-102 oxygen atom that is further hydrogen bonded to His-57, a residue which takes an active role in the catalytic cleavage event.

Asp-102 of alpha lytic protease is the recipient of a total of four hydrogen bonds: one from each of the amide

groups of Gly-56 and His-57, as well as one each from the side chains of Ser-214 and His-57 (Figure 39). In this respect alpha lytic protease is very similar to SGPA, SGPB (Figures 19a and 24) and the pancreatic serine proteases (Birktoft and Blow, 1972; Bode and Schwager, 1975; Sawyer et al., 1978; see Figures 19b and 40). Clearly, the side chain of Asp-102 is in a polar environment, albeit this residue is isolated from direct solvent contact.

The side chains of His-57 and Ser-195 in both alpha lytic protease and SGPA also deserve special consideration. It has long been proposed that the Asp-102 to His-57 couple acted, via a hydrogen bond, to activate the gamma oxygen of Ser-195, thereby conferring on this atom an abnormally enhanced nucleophilicity. It is clear from Figures 38 and 39 that in alpha lytic protease the gamma oxygen of Ser-195 is not in a favorable position for hydrogen bonding to the NE2 nitrogen atom of His-57. This distance in alpha lytic protease is long (present set of atomic coordinates gives 3.3 angstroms) and the approach is decidedly distorted. A similarly distorted interaction has been described for SGPA, SGPB and the pancreatic enzymes. These studies support the proposals of Matthews et al. (1977), that there is little, if any, interaction between the gamma oxygen of Ser-195 and the NE2 nitrogen atom of His-57. Indeed, only a weak hydrogen bonding interaction between His-57 and Ser-195 in alpha-chymotrypsin has been proposed on the basis of NMR studies (Robillard and Shulman, 1974b).

Interpretation of the native electron density map of alpha lytic protease indicates that Ser-195 has a chi1 value of -56°. This value compares well with -80° found for SGPA and -97° found for SGPB. The observed spread in chi1 values is probably a reflection of the different crystallization conditions used and from interpreting multiple isomorphous replacement maps at only 2.8 angstrom resolution. These chi1 values are also consistent with the results found for the pancreatic serine proteases, with the exception of the 'up' position assigned to the Ser-195 side chain of alpha-chymotrypsin.

K. Substrate Recognition Sites

A systematic comparison of esterase activities (Kaplan and Whitaker, 1969; Kaplan et al., 1970) and the analysis of oxidized insulin chain cleavage patterns (Whitaker et al., 1965b) have shown that alpha lytic protease preferentially cleaves on the carbonyl side of small neutral L-amino acids. Cptimal cleavage occurred at alanine residues. Alpha lytic protease also shows a marked preference for longer substrates, suggesting the possibility of additional well developed secondary binding subsites further removed from the scissile bond than the primary specificity site (Whitaker et al., 1965b).

Even though the overall sequence and tertiary structural homology of alpha lytic protease, SGPA and SGPB is high, their primary specificities are quite different.

SGPA and SGPB perform peptide bond cleavage on the carbonyl side of large hydrophobic amino acids such as phenylalanine, tyrosine and leucine (Bauer, 1976a,1978). The S1 binding sites of SGPA and SGPB, based on comparisons with alpha-chymotrypsin, are constructed from three segments of polypeptide chain: residues 191-192B, 213-219 and 224-227 (Figures 19a and 24).

It is in portions of polypeptide chain, from which the primary specificity pocket of both SGPA and SGPB are constructed, that there are non-conservative amino acid replacements and insertions in the sequence of alpha lytic protease (Table 1). Since ultimately the structure of the S1 pocket determines the overall specificity of an enzyme, these sequence alterations can be correlated with the substantially different specificity of alpha lytic protease. The major five residue insertion at Asn-217 is clearly involved in defining the substrate specificity pocket of alpha lytic protease. A substantial reduction in the volume of this pocket relative to that of SGPA or SGPB is achieved by one of these inserted residues, Val-217A. The side chain of this residue protrudes directly into the S1 binding region (Figure 39). A second important factor in decreasing the size of the primary binding pocket of alpha lytic protease is Met-192. This residue is an alanine in SGPA and SGPB. As can be seen in Figure 39, the side chain of Met-192 extends into the specificity pocket region below Val-217A.

Clearly, the placement of Met-192, coupled with the

five residue insertion at Asn-217, considerably reduces the size of the S1 binding pocket of alpha lytic protease. Indeed, model building experiments based on inhibitor studies of SGPA, to be discussed in the following chapter, indicate the primary specificity pocket of alpha lytic protease can only accommodate side chains as large as that of valine. This is in excellent agreement with observed kinetic results (Kaplan et al., 1970). Note also, that the environment of this pocket is hydrophobic, thus limiting the type of bound substrate side chains to the smaller aliphatic amino acids.

Other insertions and amino acid replacements in the sequence of alpha lytic protease in this region (Table 1) appear to be in response to the two alterations already discussed. These additional sequence changes do not directly influence the size of the primary specificity pocket. The side chain of Arg-192B, which in the native structure is found in the active site near a bound sulfate anion (Figure 38 and 39), would not be expected to limit the size of the specificity pocket in solution due to its surface position and rotational flexibility.

One portion of polypeptide chain, which forms the outer wall of the primary specificity site in alpha lytic protease (residues 214-216), is highly conserved in both SGPA and SGPB (Figures 39, 19a and 24). These residues have been shown to be responsible for binding the polypeptide backbone of a bound inhibitor in other serine proteases (Segal et

al., 1971; Sweet et al., 1974; Ruhlmann et al., 1973). Based on these structural similarities, it is likely that alpha lytic protease binds the polypeptide backbone of a substrate to its surface in a similar manner.

It is interesting to note, that the microbial serine proteases achieve alternative substrate specificities in a manner distinct from the pancreatic serine proteases. For example, alpha-chymotrypsin and elastase have specificities that are as different as those exisiting between SGPA and alpha lytic protease. Nevertheless, there are no large insertions in the sequence of elastase in the active site region (Table 1). As Figure 19b shows, alpha-chymotrypsin has a large hydrophobic pocket capable of binding a tryptophan ring. However, reference to Figure 40 shows this cavity in elastase is effectively blocked by the replacement of a valine residue for a glycine residue at position 216 and a threonine residue for a glycine residue at position 226 (Shotton and Watson, 1970). The resultant primary specificity pocket of elastase can only accommodate small neutral amino acid side chains. It is possible that with the resolution of further structures of serine proteases of elastase-like specificity, the modes of achieving this cleavage specificity, as exemplified in alpha lytic protease and elastase, will prove to be representative of the microbial and mammalian classes of these enzymes.

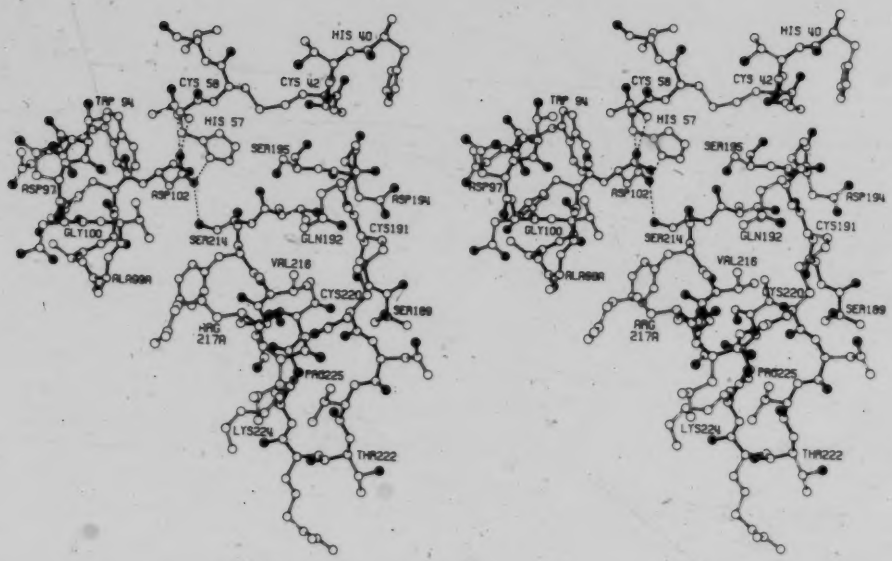


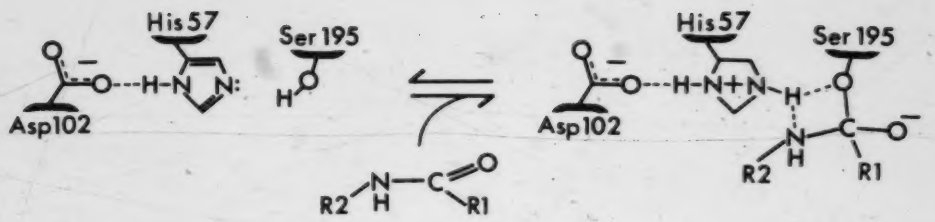
Fig. 40. Stereo-drawing of the active site of elastase (coordinates from Sawyer et al. (1978)). Two residues block access to the chymotryptic-like primary specificity pocket of elastase, allowing only small substrate side chains to be bound in this site. One of these, Val-216, is labelled in this drawing. The other, Thr-226, is situated below Val-216 in the view presented.

L. Environment of Asp-102 And Interpretation Of NMR Data

A controversy surrounds the assignment of pKa's for His-57 and Asp-102 in the active sites of serine proteases. Alpha lytic protease has played a central role in this controversy by virtue of having a single histidine in its sequence, which has allowed for less ambiguous interpretations of histidine directed experiments. Further advantages for using alpha lytic protease are derived from its pH stability and the absence of significant autolysis in solution (Kaplan et al., 1970).

A variety of techniques have been used to determine the ionization properties of the catalytic residues of serine proteases. There is agreement, that in this enzyme family one of the active site residues has a pKa of approximately 7.0. Nevertheless, it is the assignment of this pKa to a specific residue (His-57 or Asp-102) which is in dispute. The interpretation of the results of studies using alpha lytic protease to clarify this problem has necessitated the assumption that this bacterial enzyme has an active site which is homologous to the pancreatic enzymes. The present crystallographic study and the preceding structural description, convincingly demonstrates that such assumptions are valid.

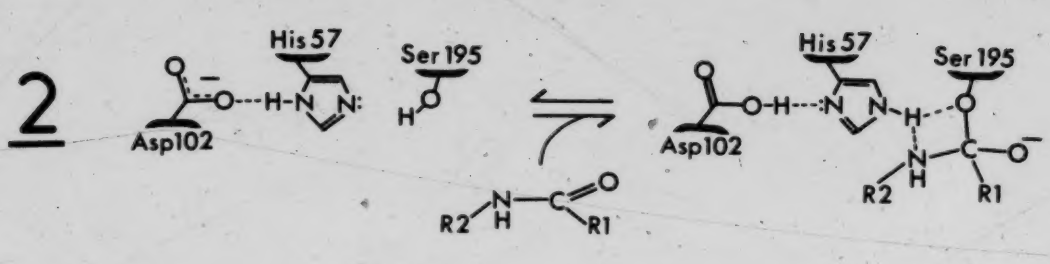
Clearly, the assignment of pKa's to His-57 and Asp-102 will have important implications on the ultimate description of the catalytic mechanism of the serine proteases. However, NMR studies have lead to two conflicting views surrounding the assignment of pka's. The first of these (Robillard and Shulman, 1974a,b) assigns each active site residue an approximately normal pKa (His-57 pKa 7.0; Asp-102 pKa 4.5). This implies that in the native enzyme and throughout the process of catalysis (optimal pH > 7.0) the proton involved in a hydrogen bond between Asp-102 and His-57 resides on His-57 (equation 1). The function of Asp-102, is to orient the imidazole ring of His-57 optimally throughout the reaction. Also implicit in this assignment is the fact that upon substrate attack and in the subsequent transfer of a

1 

proton from Ser-195 to His-57, at some point a positively charged imidazolium ring would be formed leading to a charge separation in the His-Asp couple as shown in equation 1.

The second theory, propounded by Hunkapiller et al. (1973), in effect exchanges the pKa's of the two active site residues (His-57 pKa 3.3; Asp-102 pKa 7.0). This reversal of pKa was suggested from the concept that the environment of Asp-102 is of a sufficiently hydrophobic nature that the pKa of this residue is raised by approximately 2 units. The imidazole ring is seen as having a dual role in this system. Firstly, it helps to separate the carboxyl group of Asp-102 from solvent, thereby ensuring the hydrophobicity of its environment. Secondly, by virtue of the amphoteric nature of the imidazole ring, it provides a mechanism for the relay of the net transfer of a proton from the hydroxyl group of Ser-195 to the buried basic carboxylate anion as shown in equation 2. Thus, this proposal, which is similar to the charge relay mechanism of Blow (1976), would reduce charge separation in the His-Asp couple during catalysis.

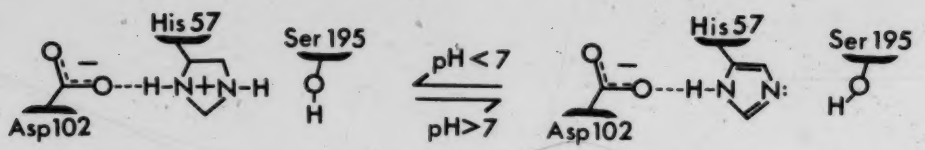
The NMR technique has held out the possibility of determining which of the two equations discussed above is

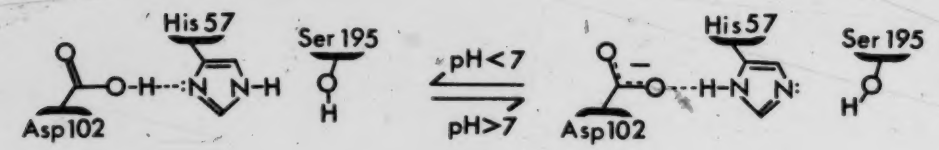
2 

correct. These studies have concentrated on proving or disproving the pH dependent processes outlined in equations 3 and 4. Clearly, if equation 3 can be shown to be correct, then His-57 and Asp-102 have normal pKa's. Alternatively, confirmation of equation 4 could be taken as proof that the pKa's of His-57 and Asp-102 were actually reversed from these normal values.

One of the first systematic NMR studies of serine proteases (Robillard and Shulman, 1972,1974a,b), among them alpha lytic protease, showed that a single broad resonance (deuterium exchangeable) could be observed at unusually low fields. This resonance shifted upfield with increasing pH and titrated with a pKa of approximately 7. This result does not unambiguously distinguish between a proton attached to a carboxyl or to an imidazole group. However, these investigators concluded, on the basis of pKa values of imidazole rings in other proteins that His-57 is normal, thus supporting equations 1 and 3.

Hunkapiller et al. (1973) approached this problem in a more novel manner. These investigators used a sample of alpha lytic protease which had been specifically enriched

3 

4 

with ^{13}C at the CE1 carbon atom of the imidazole ring of the single histidine (His-57) of this enzyme. Subsequent NMR analysis of the pH profile of the ^{13}C resonance showed it to titrate with a pKa of approximately 6.7. Rather than take the interpretation of a normal pKa for His-57, Hunkapiller and co-workers concluded that Asp-102 had the pKa of approximately 6.7. This interpretation for the ^{13}C-enriched alpha lytic protease spectra was based on two observations. Firstly, the carbon-hydrogen coupling constant for the CE1 carbon atom of the His-57 imidazole ring had a value characteristic of a neutral imidazole ring from pH 5.2 to 8.2. This suggested that protonation does not occur until much lower pH values. Secondly, in the discussion section of their paper, Hunkapiller and co-workers stated that the environment of Asp-102 is hydrophobic, so hydrophobic in fact, that the pKa of Asp-102 is likely raised to 6.7 and that of His-57 lowered to 3.3. Thus, these authors are the main proponents of equations 2 and 4.

None of the crystallographically determined enzyme structures of pancreatic and pancreatic-like microbial serine proteases solved to this point, support the premise of Hunkapiller and co-workers that Asp-102 lies in a hydrophobic environment. In each of these enzymes Asp-102 is the recipient of four highly conserved hydrogen bonds (see Figures 19a, 19b, 24, 40; Bode and Schwager, 1975). In this regard, as the present crystallographic analysis has shown, alpha lytic protease has Asp-102 in a similar disposition (Figure 39). Clearly, Asp-102 is found in a strongly polar environment, albeit isolated from direct interaction with solvent by His-57, Phe-94 and segments of the methionine loop. The structures of alpha lytic protease, SGPA and SGPB, also show that the internal hydrophilic region surrounding Asp-102 has been highly conserved during the evolution of serine proteases. This indicates that an isolated, controlled polar environment about Asp-102 is a fundamental requirement in the catalytic event. Thus, the premise that Asp-102 is in a hydrophobic environment is incorrect.

It has also been pointed out (Egan et al., 1977; Markley and Ibanez, 1978) that the coupling constant measured by Hunkapiller et al. (1973) could be in error. This measurement was hampered by the background of natural abundances and by the large widths of the resonance lines (approximately 30Hz) compared to the small change reported in the coupling constant (13Hz). Therefore, it has not been clearly established that the titration curve observed for

alpha lytic protease represents anything other than that of a normal histidine residue.

A recent NMR study of alpha lytic protease is more in agreement with the present crystallographically determined structure of this enzyme. Using an auxotroph of Myxobacter 495, Bachovchin and Roberts (1978) synthesized ^{15}N enriched alpha lytic protease at the ND1 position of the side chain of His-57. In a similar way, these authors also prepared alpha lytic protease which was ^{15}N enriched at both the ND1 and NE2 positions of the imidazole ring of His-57. This study was able to show that a proton is titrated specifically at the NE2 position of His-57 with a pKa of 7.0. Furthermore, it was shown that the tautomer of His-57 in which a hydrogen atom is bound at the ND1 position of the imidazole ring is favored. This is probably a result of the strong hydrogen bonding formed to the carboxylate anion of Asp-102. Throughout the range of pH which was monitored, it was also shown that the hydrogen atom bound to the ND1 nitrogen atom of His-57 remained attached to that atom. On this basis, it was concluded His-57 and Asp-102 have normal pKa's. Thus, both this study and most of the earlier studies, as well as the present crystallographic analysis of alpha lytic protease, indicate that equations 1 and 3 represent the true state of catalytic residues in the active site.

VI. Complexation Of A Tetrapeptide Aldehyde In The Active Site Of SGPA

A. Peptide Aldehyde Substrate Analogs

Peptide aldehyde analogs of good substrates have been shown to be very strong competitive inhibitors of serine and cysteine proteases (Westerik and Wolfenden, 1972; Ito et al., 1972; Thompson, 1973; Breaux and Bender, 1975; Hunkapiller et al, 1975; Clark et al., 1977). Such peptide analogs have an aldehyde functional group that replaces the terminal alpha-carboxyl group of a normal peptide. Under similar conditions, specific peptide aldehydes bind much more strongly to these enzymes than other analogous peptide inhibitors and substrates. This tight complexation phenomenon and the ability of aldehydes to form tetrahedral hemiacetal adducts in aqueous solutions (Lewis and Wolfenden, 1977), suggests that the complex formed by a peptide aldehyde with a serine or cysteine protease is covalent. The stability of such a covalent tetrahedral hemiacetal adduct can be viewed as a consequence of its similarity to tetrahedral intermediates postulated to occur during normal peptide hydrolysis (Wolfenden, 1972, 1976; Lienhard, 1973).

A specific peptide aldehyde has also been shown to bind tightly to SGPA. Inhibition of SGPA catalyzed hydrolysis of Ac-Pro-Ala-Pro-Phe-NH2 (pH 9.0) and of Ac-Pro-Ala-Pro-Phe-ME (pH 4.0) by Ac-Pro-Ala-Pro-Phe-al (pH 4.0 and 9.0)

and Ac-Pro-Ala-Pro-Phe-NH2 (pH 4.0) has been investigated (C.-A. Bauer, personal communication). The data show inhibition of SGPA catalyzed hydrolysis reactions by Ac-Pro-Ala-Pro-Phe-al is reversible and competitive. The aldehyde binds approximately 10^4 fold tighter ($K_i=5.0 \times 10^{-8}M$) than the corresponding peptide amide ($K_m=5.4 \times 10^{-4}M$) at an optimal pH (9.0). On going from pH 9.0 to 4.0, there is an approximate 2 fold decrease in binding of the peptide amide ($K_i=1.2 \times 10^{-3}M$ at pH 4.0). The binding of the peptide aldehyde is clearly much more pH dependent ($K_i=2.0 \times 10^{-6}M$), there being a 40 fold decrease in binding upon lowering the pH to 4.0. Even so, the peptide aldehyde is a very effective inhibitor, binding about 600 fold tighter to SGPA than the peptide amide at this acid pH. This tight binding of a specific peptide aldehyde to SGPA, both at pH 4.0 and 9.0, is comparable to the binding of specific peptide aldehydes to elastase (Thompson, 1973).

In order to more clearly establish the nature of peptide aldehyde complexation in the active sites of serine proteases, 2.8 angstrom resolution crystallographic data from the complex of SGPA and the specific peptide aldehyde Ac-Pro-Ala-Pro-Phe-al has been collected. As indicated by its chemical formula, this peptide aldehyde has a C-terminal phenylalanine residue, which has an aldehyde function in place of a carboxyl group. The selection of this peptide aldehyde for study, was governed by the primary specificity of SGPA for Phe, Tyr and Leu residues in subsite S1. In

addition, earlier studies had suggested peptides incorporating the sequence Pro-Ala-Pro (P4-P2), bind in only cne mode on the surface of SGPA (Bauer et al., 1976a). The peptide aldehyde used in this study was generously supplied by Drs. C.-A. Bauer and R.C. Thompson.

B. Crystallographic Data Collection

A suitable crystal of the SGPA - peptide aldehyde complex was prepared by immersing a native SGPA crystal into an approximately 1mM peptide aldehyde solution containing 1.5M NaH(2)PO(4) at pH 4.1. The rate of peptide aldehyde penetration could be conveniently followed as a function of the change in the refractive index of the crystal as monitored under cross-polarized light in a petrographic microscope. Upon completion of this soaking procedure (6hr), 2.8 angstrom resolution X-ray diffractometer data were collected, processed and scaled as previously described for native SGPA crystals. Relevant crystallographic data collection and processing statistics are shown in Table 26. Structure factor amplitude differences observed between the diffraction data of the SGPA - peptide aldehyde complex and of native SGPA were used tc compute a difference electron density map. The coefficients for this Fourier synthesis were the figure of merit weighted differences, {F(P+I) - F(P)}, where F(P) represents the structure factor amplitudes of native SGPA and F(P+I) are those amplitudes from the crystal of the aldehyde complex. The phases and figure of

TABLE 26

SGPA/Peptide Aldehyde Complex Diffraction Data Statistics

Data	Native Enzyme	Complex
a = b, c (angstroms)	55.14, 54.81	55.17, 54.62
No. of reflections measured	9165	4582
Max. absorption correction (%)	14.9	6.7
Max. crystal decay (%)	10.2	5.3
[1]R(sym) (%)	1.7	1.5
No. reflections merged	4113	304
Percent reflections (I > 3sigma(I))	96.2	94.5
Absolute scale	10.78	9.48
Overall isotropic B (angstroms)^2	12.4	13.0
[2]R(I) (%)	-	10.0

[1]R(sym) is defined in Table 5.
[2]R(I) is calculated in the same manner as R(D) (Table 7), using inhibitor complex rather than heavy-atom derivative amplitudes.

merits for this difference map were obtained from the original multiple isomorphous replacement protein phase determination of native SGPA.

The three-dimensional interpretation of the difference electron density map was carried out in a Richards optical comparator (Richards, 1968) by fitting Kendrew-Watson skeletal units connected to depict the chemical sequence of the bound aldehyde. Coordinates for all non-hydrogen peptide aldehyde atoms were measured from the resultant model and

fitted using Diamond's (1966, 1974) model building procedure. The overall r.m.s. deviation of the model built structure from the measured coordinates was 0.11 angstroms. A final fit of the difference map was accomplished using the MMS-X graphics system. The final coordinates for the SGPA/peptide aldehyde complex are given in Appendix 2. These coordinates are in orthogonal angstrom units which correspond to the crystallographic unit cell of SGPA.

C. Difference Map Interpretation

The difference electron density map in the region of the active site of SGPA is illustrated in Figure 41, with the interpretation of that electron density shown in terms of a superimposed model. As Figure 41 shows, the main details of the difference electron density are explained by the model of the bound tetrapeptide aldehyde that has been adjusted to fit this difference density. In addition, this difference map shows that there is a major conformational change in the SGPA molecule which takes place on forming the aldehyde complex. This change is seen as the large positive and negative peaks just to the left (in this Figure) of the main portion of the difference electron density representing the bound peptide aldehyde. The position of the His-57 side chain in native SGPA is represented by the negative peak labelled H(n) in Figure 41. This side chain moves to a new position in the aldehyde complex which is coincident with the positive peak labelled H(c). Clearly a major disruption

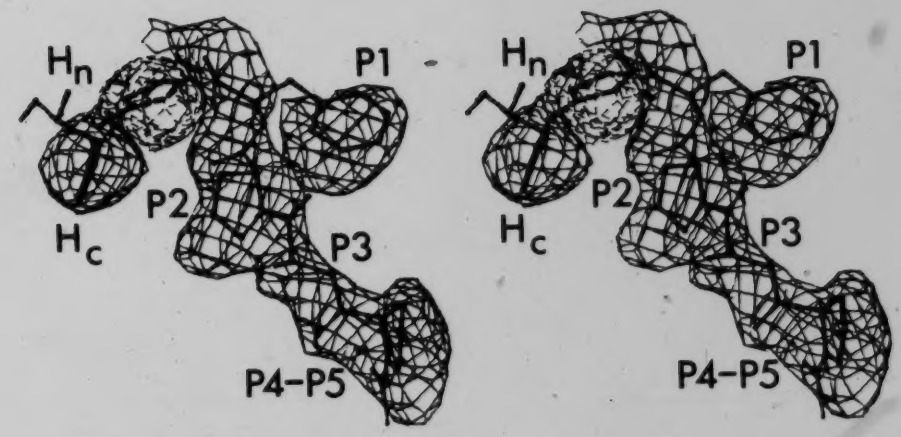


Fig. 41. A stereo-representation of the difference electron density map of the SGPA/peptide aldehyde complex at 2.8 angstrom resolution. Both the positive (solid thin lines, +0.066e/(angstroms)^3) and negative (dashed thin lines, -0.066e/(angstroms)^3) difference electron density envelopes in the immediate active site region of SGPA are presented. The standard error of this map was estimated to be 0.029e/(angstroms)^3 (Henderson and Moffat, 1971), with the highest difference electron density peak being 13 sigma above background. Superimposed on this drawing is the fitted molecular model of the bound peptide aldehyde. Also shown, is the position of the side chain of His-57 in native SGPA (heavy dashed lines, labelled H(n)) and this same residue side chain in the inhibitor complex (heavy solid lines, labelled H(c)). Only those atoms beyond the beta-carbon of His-57 move upon peptide aldehyde binding.

of this active site residue has taken place.

Another important feature of the difference map of Figure 41 is that the difference electron density associated with the bound peptide aldehyde is continuous with native electron density associated with the gamma oxygen atom position of the active site residue Ser-195. This is more clearly shown in Figure 42, which shows a portion of the figure of merit weighted, {2F(P+I)-F(P)} exp(i alpha(P)) electron density map in the region of the gamma oxygen atom

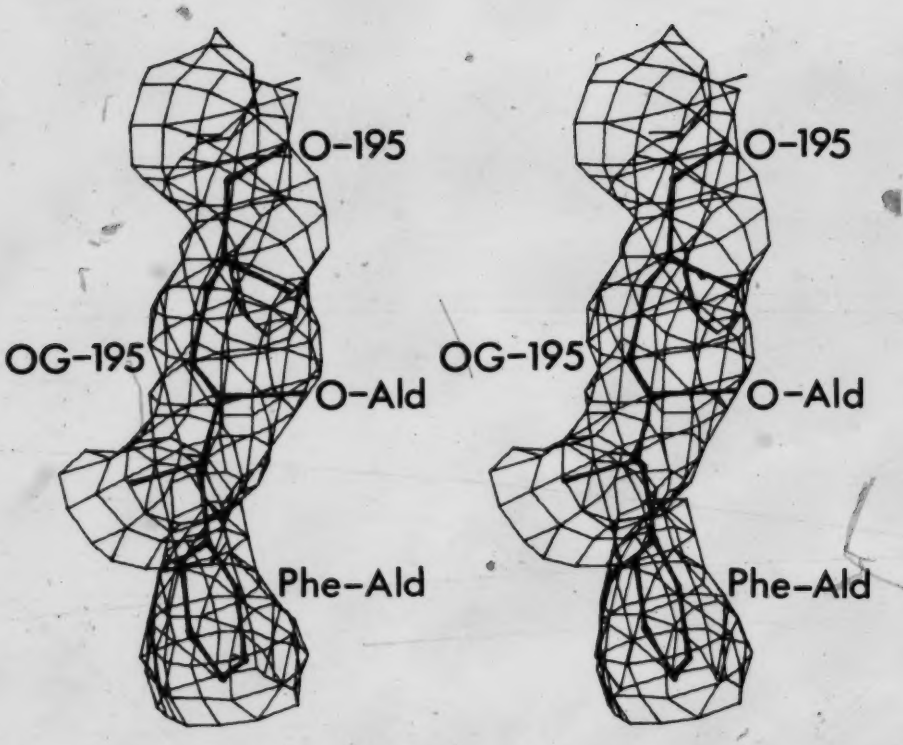


Fig. 42. Stereo-representation of a portion of the SGPA/peptide aldehyde complex electron density map in the region of the covalent bond between the active site residue Ser-195, and the P1 phenylalaninal residue of the bound aldehyde inhibitor. The contour envelope is drawn at 0.34e/(angstroms)^3. The atomic model, consisting of the side chain of Ser-195 and the P1 residue of the bound inhibitor, is drawn with thick solid lines.

of Ser-195 and the aldehyde carbonyl carbon atom. This map represents the approximate overall electron density distribution of the SGPA/peptide aldehyde complex (Blundell and Johnson, 1976). The continuous electron density observed in Figure 42 confirms the covalent nature of the linkage between this specific peptide aldehyde and SGPA.

There are also a number of smaller positive and negative peaks in the difference electron density map (these

are not shown in Figure 41). The positive peaks can be attributed to new positions of bound solvent (i.e. water molecules) in the aldehyde crystals, whereas the negative peaks represent the positions of solvent molecules which have been displaced from the native enzyme upon complex formation with the aldehyde inhibitor.

A stereo-drawing of the 2.8 angstrom resolution interpretation of the peptide aldehyde bound in the active site region of SGPA is shown in Figure 43. Comparison of Figure 43, with a similar drawing of only the active site of SGPA (Figure 19a), shows there are a number of intermolecular contacts involving both hydrogen bonding and van der Waals' interactions formed by the bound peptide aldehyde in addition to the covalent bond to Ser-195.

The primary specificity pocket of SGPA, as delineated by the phenylalanine side chain of the bound peptide aldehyde, is defined by three regions of the enzyme: the polypeptide chain from Ala-192 to Pro-192B, the polypeptide chain from Ser-214 to Gly-218, and residue Thr-226. As can be seen in Figure 43, the primary specificity pocket region is generally hydrophobic in character. The phenyl ring of P1 phenylalaninal is aligned in subsite S1 so that it is approximately coplanar with the planes of peptide groups lining both sides of this binding site. Indeed, this may be the only allowed conformation, as the S1 subsite is not sufficiently large enough to accommodate an aromatic ring in alternative conformations (Figure 43). The P2 proline

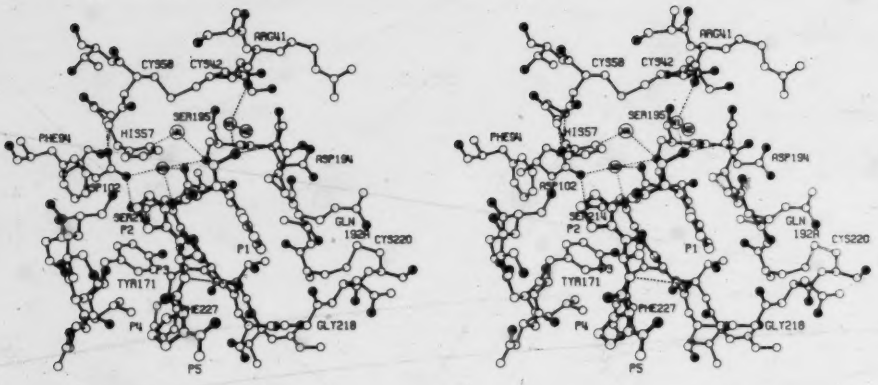


Fig. 43. Stereo-drawing of the SGPA/peptide aldehyde complex as determined from the 2.8 angstrom resolution difference electron density map. Polypeptide main chain bonding of the enzyme is indicated with filled black bonds, as are all interatomic bonds of the bound peptide aldehyde. All oxygen atoms are distinguished by filled black circles. Hydrogen bonds to active site residues and those formed to the bound inhibitor are shown as dashed lines. The new position of the side chain His-57 in the complex is also drawn. Solvent molecules bound upon peptide aldehyde complexation are indicated by the symbols (W1-W4). Potential hydrogen bonds formed with solvent molecules are also indicated by dashed lines.

residue of the inhibitor is bound in a second surface depression, which forms the S2 binding subsite. This binding pocket is also hydrophobic in nature, being formed by the side chain of Tyr-171 and portions of main chain in the vicinity of Ser-174. The side chain of His-57 in its native conformation (Figure 19a) would also form a portion of the S2 subsite. In the present inhibitor complex this side chain is rotated to lie above (in Figure 43) the P2 proline side chain. As shown in Figure 43, the side chain of P3 alanine forms no interactions with the enzyme surface and is oriented into surrounding solvent. The terminal

N-acetyl-prolyl moiety (P5-P4) of the aldehyde inhibitor lies in a poorly defined binding site at the extremity of the substrate binding region. Nevertheless, this portion of the inhibitor is well resolved in the difference electron density map (Figure 41). Major contacts between the enzyme surface and the inhibitor in the S5-S4 binding subsites involve the side chains of Val-169, Tyr-171 and parts of the main polypeptide chain at residues Gly-216 and Ser-217.

There are five hydrogen bonds formed between SGPA and the aldehyde inhibitor, all of which involve main chain polypeptide atoms of SGPA. Three of these hydrogen bonds form an approximate anti-parallel beta sheet configuration between the inhibitor P1 phenylalaninal and P3 alanine residues, and the enzyme surface residues Ser-214 and Gly-216. Formation of these hydrogen bonds is accompanied by a small reorientation of the carbonyl oxygen atoms of Ser-214 and Gly-216. These minor conformational changes are not indicated in Figure 43.

Two further hydrogen bonds involve the aldehyde group oxygen atom and the two peptide amide groups of Gly-193 and Ser-195. Due to the close proximity of these hydrogen bonds to active site residues, it is probable that these interactions play an important role in correctly positioning the susceptible peptide bond of a true substrate. The present study, although that of a bound substrate analog, suggests the mode of binding taken by a substrate in the oxyanion hole (residues 193-195).

Interpretation of the electron density maps of Figures 41 and 42, places the aldehyde carbon atom within covalent bond distance of the gamma oxygen atom of Ser-195 (1.5 angstroms). No movement of the gamma oxygen atom from its native position is observed on the binding of the aldehyde. Figure 43 shows the environment of the aldehyde carbon and oxygen atoms along with the two hydrogen bonds formed in the oxyanion hole. Nearby, there are two reasonably well ordered solvent molecules (tentatively identified as water molecules). These are labeled W1 and W2 in Figure 43. One of these solvent peaks is within hydrogen bonding distance of the aldehyde oxygen atom.

In addition to the two solvent molecules, W1 and W2, the difference electron density map indicates that two additional solvent molecules, W3 and W4 (herein also assumed to be water molecules), are bound in the active site region upon peptide aldehyde complexation. Water W3 is situated near the original native position of the imidazole ring of His-57, in close proximity to Asp-102, to the carbonyl oxygen atoms of Thr-213 and Ser-214, and to the OG oxygen atom of Ser-195. It is clear from Figure 43 that the carboxylate of Asp-102 has been exposed to solvent upon the movement of the imidazole ring of His-57. The fourth water molecule W4 forms a hydrogen bond interaction with the ND1 nitrogen atom of His-57 at its new position and the OG oxygen atom of Ser-195.

The binding of this tetrapeptide aldehyde also

displaces a number of well ordered solvent molecules from the surface of SGPA. The native electron density map of SGPA (Figure 18) has a solvent molecule which coincides with the position of the CG carbon atom of the P1 phenylalaninal residue, thus explaining the weak density connecting the phenyl ring to its alpha carbon atom. A second solvent molecule is displaced near the para position of this ring. The P2 Pro residue displaces a solvent peak which is close to the main chain carbonyl oxygen atom of Ser-214. There are also a number of less well defined solvent peaks that have been displaced by the N-acetyl-prolyl moiety.

D. Comparison With Solution Data And Other Serine Proteases

Several structural studies of inhibitor binding to the pancreatic serine proteases have been completed. These include binding chloromethyl ketone peptides to gamma-chymotrypsin (Segal et al., 1971) and examining the bound conformation of naturally occurring inhibitors to bovine and porcine trypsin (Ruhlmann et al., 1973; Huber et al., 1974; Sweet et al., 1974). Comparison of these studies with the present peptide aldehyde - SGPA complex shows a number of similarities. In each case, the portion of polypeptide chain composed of residues 214 to 216, forms an approximate anti-parallel beta sheet structure with the main chain polypeptide backbone of the bound inhibitor. Furthermore, the side chains of the bound inhibitor, P1 through P3, are similarily disposed on the surface of each

enzyme. Thus, the present study confirms that SGPA binds inhibitors and therefore likely substrates, as do the pancreatic serine proteases.

A number of features of the present inhibitor binding study serve to explain solution studies carried out with SGPA. Such studies have indicated that the primary specificity site of SGPA is less well defined than its pancreatic counterpart in alpha-chymotrypsin. A comparison of the native structures of these two enzymes has already indicated that this is likely to be the case (see Figure 19). Further comparisons of the bound conformation of the P1 phenylalanine residue of chloromethyl ketone peptides bound to gamma-chymotrypsin (Segal et al., 1971) with that of the P1 phenylalaninal residue of the peptide aldehyde bound to SGPA, confirms the less specific nature of binding interactions in the primary specificity pocket of SGPA. Also, while the primary specificity pocket of alpha-chymotrypsin is sufficiently deep to accommodate a tryptophan side chain, this same site in SGPA is more shallow. Indeed, as Figure 43 shows, even the side chain of tyrosine is likely to fit tightly in the S1 subsite of SGPA. The smaller size of the primary specificity pocket of SGPA, explains this enzyme's poor ability to cleave peptide bonds at tryptophan residues (Bauer, 1978).

Solution studies of SGPA have also suggested the presence of additional well defined subsites further removed from the scissile bond. Subsites indicated to play a role in

binding substrates include S3' through to S4. Although the present study can not be used to identify binding subsites C-terminal to the scissile bond, those of S1 through S5 can be discerned. As shown in Figure 43, subsite S2 is rather well defined. Subsite S3, while not having a binding pocket region, does form two hydrogen bonds with the main chain carbonyl and amide groups of the P3 alanine residue. The S4 subsite is poorly defined, but can still be considered as a shallow hydrophobic depression. As Figure 43 shows, there is little indication of further subsites beyond S4 that would form specific contacts with a bound substrate.

Peptides used in previous solution studies and in the present investigations, have been designed around the knowledge that proline residues bind poorly in the S3 subsite (Bauer et al., 1976a). Thus, by placing a proline residue at the P2 and P4 positions, a single mode of peptide binding is more likely. The present crystallographic study suggests two possible reasons for the poor binding capacity of the S3 binding subsite for proline residues. Firstly, a P3 proline residue could not form both of the hydrogen bonds that the P3 alanine residue of the present study does, since the imino nitrogen of proline is unavailble for such an interaction. Secondly, model building studies indicate the constrained conformation of a proline residue would direct the remaining N-terminal portion of a bound peptide off the surface of the enzyme. This would result in the loss of the binding energy resident in the S4 subsite, which has been

shown to make a substantial contribution in substrate and inhibitor binding (Bauer et al., 1976a,b).

In contrast to the secondary binding subsites of SGPA, those of gamma-chymotrypsin are poorly developed (Segal et al., 1971,1972). Indeed, only Ile-99 of gamma-chymotrypsin is expected to play a role in S2 interactions. This is consistent with solution studies showing that alpha-chymotrypsin is less dependent on interactions remote from the scissile bond during peptide catalysis (Bauer et al., 1976a,b). The more specific nature of secondary subsites in SGPA arises from the extended conformation of the methionine loop in SGPA (see Figures 14a and 19a). This loop is more compact and positioned out of the active site region in in alpha-chymotrypsin (see Figures 14b and 19b). As suggested by Bauer (1978), the specificity shown by the secondary binding subsites of SGPA, may compensate for the lack of binding specificity shown by this enzyme in the primary specificity pocket.

E. Peptide Aldehydes As Transistion State Analogs

The peptide aldehyde used in this study binds to SGPA several orders of magnitude more tightly than the corresponding peptide amide. According to transisition state theory (Wolfenden, 1972), this is consistent with the peptide aldehyde - SGPA complex being an analog of the peptide amide - SGPA transition state intermediate. The binding modes of both non-specific and specific peptide

aldehydes to serine proteases have been discussed by several authors (Thompson, 1973; Breaux and Bender, 1975; Schultz and Cheerva, 1975; Gorenstein et al., 1976; Lowe and Nurse, 1977; Chen et al., 1979). In summary, two alternative binding modes have been suggested: (1) covalent bond formation between the aldehyde carbon atom and the active site Ser-195 gamma oxygen atom, or (2) non-covalent interactions between the peptide aldehyde (or hydrate) and the active site of the enzyme.

The x-ray crystallographic results presented above represent the first structural determination of an enzyme - peptide aldehyde complex. These results (as depicted in Figure 42) present direct evidence that a stable covalent tetrahedral hemiacetal adduct is formed by a specific peptide aldehyde in the active site of a serine protease. Indeed, inspection of the difference electron density map of this complex (Figure 41) and its subsequent interpretation places the aldehyde carbon atom of the bound peptide approximately 1.5 angstroms from the gamma oxygen atom of Ser-195 in the active site of SGPA. Figure 44, in a stylized fashion, summarizes the covalent and other non-covalent peptide aldehyde - enzyme interactions found.

Another important feature of the SGPA - peptide aldehyde complex is the reorientation of the side chain of the active site residue His-57. This is achieved via two rotations: one of approximately 90° about its alpha-carbon to beta-carbon bond and another of approximately 30° about

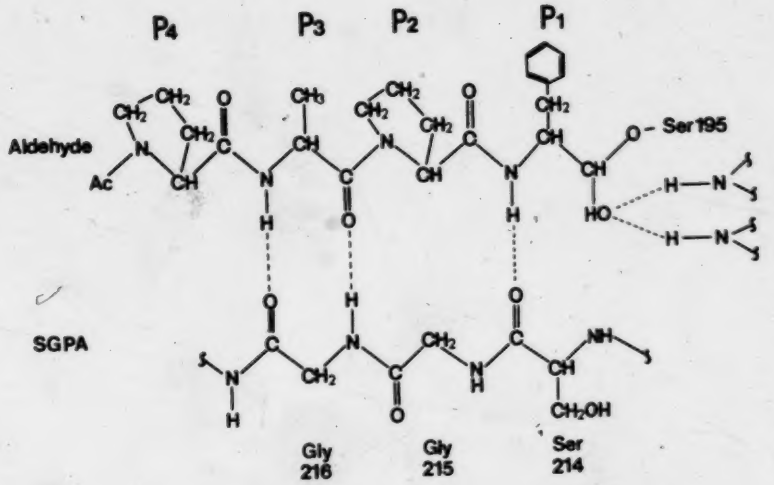


Fig. 44. Stylized drawing of the interactions formed between the bound peptide aldehyde and SGPA. Along with the covalent bond formed to the gamma oxygen atom of Ser-195, five hydrogen bonds are formed with surface enzyme groups. Three of these hydrogen bonds take on an approximate anti-parallel beta sheet conformation.

the beta-carbon to gamma-carbon bond. The possibility of His-57 rearrangement in the active sites of serine proteases upon peptide aldehyde complexation has been postulated in a paper describing an NMR study of alpha lytic protease (Hunkapiller et al., 1975).

As discussed earlier (Chapter 5) alpha lytic protease is highly homologous to SGPA in structure, particularly about active site residues (See Figures 19a, 29 and 39). Like SGPA, alpha lytic protease binds a specific peptide aldehyde much tighter than the corresponding peptide alcohol and methyl ester (Hunkapiller et al., 1975). NMR analysis of the complex formed by alpha lytic protease and the peptide aldehyde Ac-Ala-Pro-Ala-al below pH 5.0 suggests that a

hemiacetal tetrahedral addition complex is formed and the presence of a protonated imidazole cation with considerable mobility is detected. The particular histidine involved is His-57, the only such residue in the polypeptide sequence. These results are directly comparable to the expulsion of His-57 into solvent in the SGPA - peptide aldehyde complex at pH 4.1. As has been observed for native SGPA, further NMR analysis of native alpha lytic protease between pH 3.5 and 6.0, indicates His-57 remains tightly lodged in the active site. Clearly, for the alpha lytic protease - inhibitor complex, His-57 mobility is a function of peptide aldehyde complexation under pH 5.0.

NMR analysis of the alpha lytic protease -peptide aldehyde complex above pH 7.0 indicates His-57 retains its native conformation upon hemiacetal formation (Hunkapiller et al., 1975). These results suggest a reorientation of the tetrahedral hemiacetal addition complex takes place upon lowering pH, which is responsible for the movement of His-57 out of the active site. Such an explanation is consistent with the present SGPA -peptide aldehyde complex at pH 4.1 (Figure 41). In this complex the bound aldehyde carbon atom is positioned in close contact with the native position of the side chain of His-57. The aldehyde hydrogen is oriented even closer by the solvated position of the aldehyde oxygen atom in the oxyanion hole. If the model is adjusted so that the aldehyde oxygen atom is optimally positioned in the oxyanion hole, as would be expected at pH > 7.0, where SGPA

is most catalytically active, these close contacts are relieved. A similar movement of the His-57 side chain of SGPB has also been observed for steric reasons resulting from a covalently bound pipsyl group on Ser-195 (Codding et al., 1974).

Nevertheless, as will be discussed in the following inhibitor study, not all peptide aldehydes bound to SGPA at low pH result in a movement of His-57. Indeed, it seems this feature of peptide aldehyde binding is dependent on the amino acid composition of the bound inhibitor.

VII. The Complex Formed By Chymostatin And SGPA

A. Chymostatin: Chemical Structure And Specificity

Chymostatin is a naturally occurring inhibitor of serine proteases that have chymotryptic specificity (Umezawa et al., 1970; Feinstein et al., 1976) and is isolated from the culture filtrates of a variety of Streptomyces species (Umezawa, 1976). Chemical and NMR data have been interpreted to propose an unusual chemical structure, shown in Figure 45a, for this inhibitor (Tatsuta et al., 1973). The novel structural features of chymostatin include: an aldehyde function which is incorporated into the P1 residue, a cyclized arginine residue at P3 and an ureido-moiety linking the P3 and P4 residues, which reverses the C to N direction of the polypeptide chain at the P4 phenylalanine residue. Chymostatin, when isolated, is a mixture of three components which differ in the identity of the amino acid residue at P2 of the polypeptide chain. Chymostatin A has a leucine residue at P2 whereas chymostatin B and chymostatin C have valine and isoleucine residues, respectively (Tatsuta et al., 1973).

As was the case with the previously discussed tetrapeptide aldehyde, chymostatin is believed to be a good inhibitor of serine proteases by virtue of its aldehyde function and this group's ability to form a tetrahedral hemiacetal adduct with Ser-195. Chymostatin is an effective inhibitor of alpha-chymotrypsin, the mammalian serine

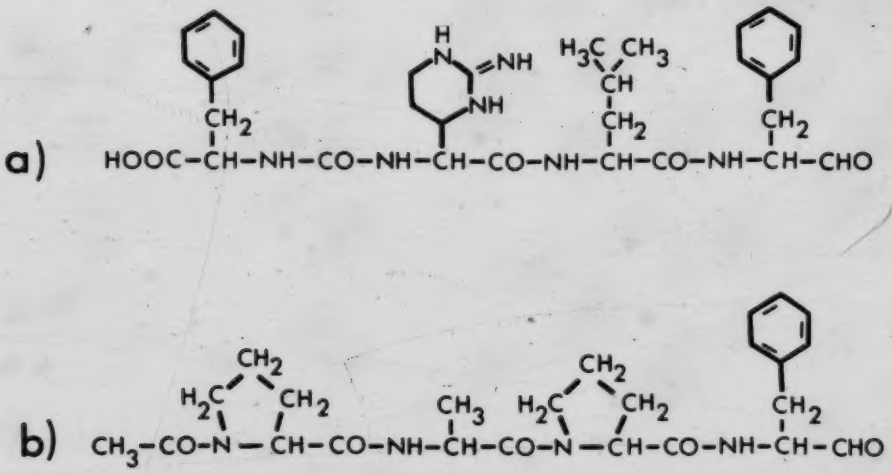


Fig. 45. A structural comparison of (a) chymostatin A, a naturally occurring peptide aldehyde of unusual chemical structure, and (b) the synthetic tetrapeptide aldehyde inhibitor used in the previous binding study of SGPA. Note the similar overall size of these two inhibitors and the presence of a terminal phenylalaninal residue in each.

protease which has a primary specificity pocket for which the P1 phenylalanyl side chain of this inhibitor is suited (Aoyagi and Umezawa, 1975). Unfortunately, similar binding studies with SGPA have not been carried out. Nevertheless, as shown in Figure 45b, the tetrapeptide aldehyde used in the previous inhibitor study of SGPA, has some characteristics in common with chymostatin. Note the similar overall size and the presence of a C-terminal phenylalaninal residue in each of these inhibitors. On this basis it would be reasonable to expect a similar overall binding mode for these two inhibitors. In order to confirm the unusual chemical structure of chymostatin, and to define more clearly the mode of its inhibitory activity, 2.8 angstrom

resolution crystallographic data from the complex formed by SGPA and chymostatin have been collected.

B. Collection Of Three-Dimensional Data

Chymostatin was purchased from Peninsula Laboratories (Lot 230715). Chemical analysis by the Protein Research Foundation (Osaka, Japan) indicated that the sample was composed of 78% chymostatin A, 17% B and 5% C; that is, with Leu, Val and Ile, at the P2 position, respectively. A saturated solution of chymostatin was prepared in 1.5M NaH(2)PO(4) at pH 4.1 and a crystal of SGPA was placed in this solution. The progress of penetration of the inhibitor solution into the native SGPA crystal was followed by monitoring the change in interference color through a polarizing microscope. This soaking procedure required three weeks to go to completion. The SGPA-chymostatin complex crystal was then mounted in a glass capillary tube and a set of diffraction data was collected and processed as described in the previous peptide aldehyde inhibitor study. The relevant crystal data statistics are given in Table 27.

As in the previous inhibitor study, a difference electron density map was calculated and a Watson-Kendrew model of chymostatin was fitted into this map using an optical comparator (Richards, 1968). The coordinates of the non-hydrogen atoms of chymostatin were then measured from this model and fitted using Diamond's (1966, 1974) model building procedure. The overall r.m.s. deviation of the

TABLE 27

Crystal Data For The SGPA/Chymostatin Complex

a = b, c (angstroms)	55.05, 54.63
Max. absorption correction (%)	10.3
Max. crystal decay (%)	13.3
[1]R(sym) (%)	1.2
No. reflections merged	283
Percent reflections (I > 3sigma(I))	99.4
Absolute scale	10.94
Overall isotropic B (angstroms)^2	11.5
[2]R(I) (%)	11.6

[1]R(sym) is defined in Table 5.
[2]R(I) is calculated in the same manner as R(D) (Table 7), using inhibitor complex rather than heavy-atom derivative amplitudes.

model built structure from the measured coordinates was 0.19 angstroms. The coordinates of the stereo-chemically fitted model were then transferred to an MMS-X graphics system and further manipulated to fit optimally in the difference electron density map. The final coordinates for the SGPA/chymostatin complex are given in Appendix 3. These coordinates are in orthogonal angstrom units corresponding to the crystallographic unit cell of SGPA.

C. Difference Electron Density Interpretation

The final model of chymostatin superimposed on its difference electron density distribution is shown in Figure 46. For clarity, negative electron density contours are not shown in this Figure. The proposed chemical model of chymostatin (Figure 45a) agrees very well with the difference electron density distribution. Thus, the present study confirms the proposed molecular structure of chymostatin (Tatsuta et al., 1973). Also shown in Figure 46, are the OG, CB and CA atoms of the side chain of Ser-195. Difference electron density representing the position of bound chymostatin is continuous with the gamma oxygen atom position of Ser-195. This indicates that a covalent bond has been formed from the carbonyl carbon atom of the aldehyde function of chymostatin to the gamma oxygen atom of Ser-195. This bond is approximately 1.5 angstroms in length. The side chain of Ser-195 retains its native position during complex formation and it should be noted that the position of His-57 is also unperturbed upon binding chymostatin.

A stereo-drawing of the 2.8 angstrom resolution fitted model of chymostatin, bound in the active site of SGPA is shown in Figure 47. In addition to the covalent bond formed to Ser-195, chymostatin forms a number of other non-covalent interactions. The aldehyde oxygen atom of the P1 phenylalaninal residue is located in the oxyanion hole, , forming two hydrogen bonds, one to each of the peptide amide groups of Gly-193 and Ser-195. The polypeptide backbone of

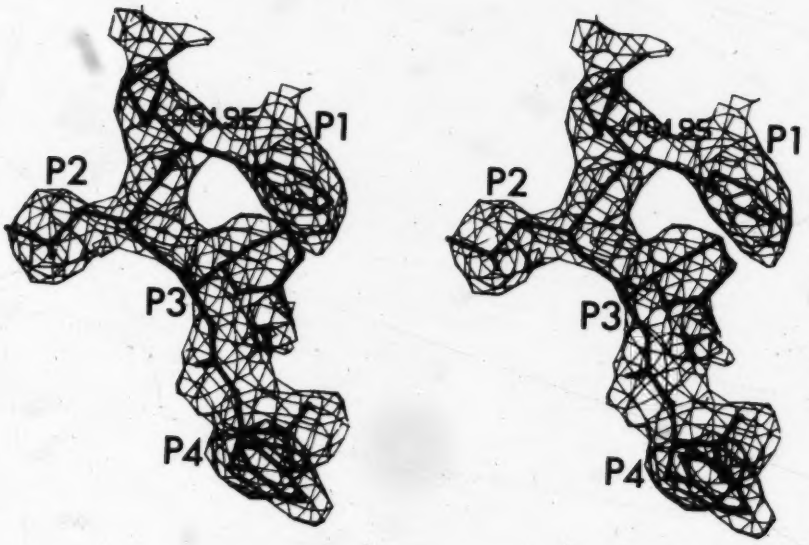


Fig. 46. Stereo-representation of the difference electron density map of the chymostatin/SGPA complex at 2.8 angstrom resolution in the region of the active site of the enzyme. The positive contour envelope shown is drawn at a level of 0.066e/(angstroms)^3. The standard error of this map was estimated to be 0.038e/(angstroms)^3 (Henderson and Moffat, 1971). The highest positive peak on this difference map is more than 10 sigma above background. Also shown superimposed, is the final fit of the molecular model of chymostatin (thick lines) to this difference electron density.

chymostatin then continues on to form an anti-parallel beta sheet interaction with residues 214 to 216. A total of three hydrogen bonds are formed in this manner. One of these is between the peptide amide group of the P1 phenylalaninal residue and the carbonyl oxygen atom of Ser-214. The remaining two hydrogen bonds are formed from the carbonyl oxygen atom and the peptide amide group of the P3 cyclized arginine residue, to the amide group and carbonyl oxygen atom of Gly-216, respectively.

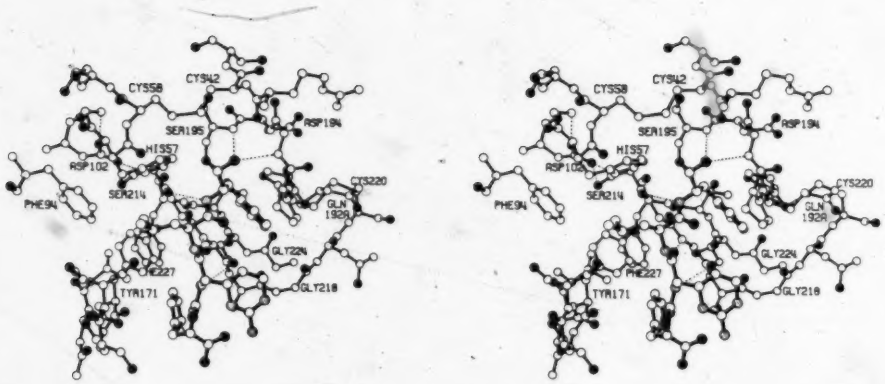


Fig. 47. Stereo-drawing of the chymostatin/SGPA complex as determined by fitting the difference electron density map of Figure 46. Only non-hydrogen atoms are illustrated. Polypeptide main chain bonding of SGPA, as well as all interatomic bonds of the inhibitor, are shown as solid black bonds (those of chymostatin are slightly wider). Hydrogen bonds formed by active site residues and those formed by the bound inhibitor to the enzyme surface, are indicated by thin dashed lines. All oxygen atoms present are denoted as solid black circles, while only the nitrogen atoms of chymostatin are represented by striped circles.

The P1 phenylalaninal side chain of chymostatin lies in the primary specificity pocket of SGPA. Interpretation of the difference electron density in the S2 secondary binding subsite, indicates there is no evidence that a valine or isoleucine residue has been bound at this site. This difference electron density, however, is fit by a leucyl side chain very well (Figure 46). Therefore, it seems that chymostatin A (P2 Leu, 78% of the mixture used) is selectively bound in the active site of SGPA, over chymostatins B and C. Nevertheless, the present experiment does not rule out the presence of minor amounts of valyl or isoleucyl side chains being bound in the S2 subsite, which

may fall below the level of detection of this study. The cyclized arginine residue of chymostatin is bound in subsite S3. As Figure 46 shows, the difference electron density map confirms the unusual cyclized nature of this side chain. The difference electron density map is also consistent with the proposed ureido-grouping of chymostatin located between residues P3 and P4. The terminal phenylalanine residue of chymostatin is also well resolved, being positioned near the extremity of the substrate binding region of SGPA (Figure 47).

D. Comparison With The Tetrapeptide Aldehyde - SGPA Complex

In many respects the binding mode of the naturally produced chymostatin and that of the synthetically constructed peptide aldehyde, Ac-Pro-Ala-Pro-Phe-al, are similar (Figures 43 and 47). Both bind covalently to the gamma oxygen of Ser-195 to form a tetrahedral hemiacetal addition complex. In each case, this bond is approximately 1.5 angstroms in length and its formation does not cause a movement in the native position of the side chain of Ser-195. Furthermore, the aldehyde oxygen atom is oriented into the oxyanion hole in each complex. Other similarities are observed in the general disposition of bound side chains of the residues that make up each inhibitor and the manner in which their respective polypeptide backbones form an anti-parallel beta structure with residues 214 to 216. Figure 48 shows a comparison of the bound conformations of

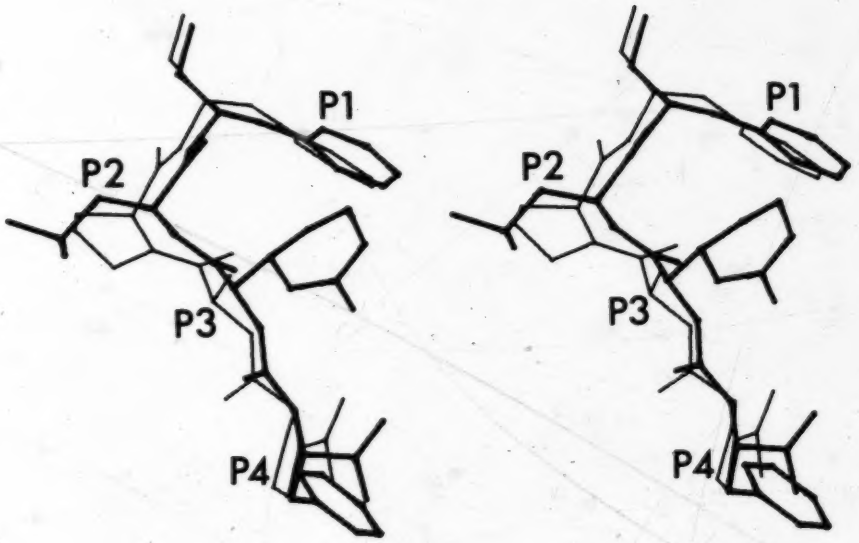


Fig. 48. Stereo-drawing of chymostatin (thick bonds) and the synthetic Ac-Pro-Ala-Pro-Phe-al inhibitor (thin bonds) superimposed. The inhibitor conformations presented were determined from fitting their respective inhibitor/SGPA difference electron density maps. The orientation of this drawing is the same as that of Figure 46.

these two inhibitors as interpreted from their respective difference electron density maps.

The two inhibitor/SGPA complexes do however, differ in one major respect. In the complex of the synthetic tetrapeptide aldehyde, the imidazole ring of His-57 is rotated away from its native position, whereas this does not occur in the chymostatin/SGPA complex. As discussed previously, His-57 movement is apparently the result of close contacts being formed with the covalently bound synthetic tetrapeptide aldehyde group. It was earlier noted that if the aldehyde group were optimally oriented in the

oxyanion hole, these close contacts could be relieved.

Also, inspection of Figure 48 shows residues P1 to P3 of chymostatin are shifted (to the right in this figure) from those of the synthetic peptide aldehyde. The overall r.m.s. deviation between non-hydrogen atoms of the polypeptide backbone for residues P1 to P3 of these inhibitors is 0.8 angstroms. The polypeptide backbone shift observed for chymostatin likely arises from the tight fit of this inhibitor's leucyl side chain in the S2 subsite. As Figure 47 shows, this side chain makes intimate contacts with Tyr-171, Phe-94, His-57 and Ser-174. To avoid prohibitive contacts in the S2 subsite, the polypeptide backbone of chymostatin is oriented further away from this subsite than is the corresponding synthetic peptide aldehyde complex.

Two factors could potentially account for the differing His-57 conformations observed in the two inhibitor/SGPA complexes investigated. Firstly, associated with the shift in bound polypeptide backbone position, is a more optimal fit of the aldehyde group of chymostatin in the oxyanion hole than is observed for this group in the previous inhibitor study. Note also, that the aldehyde oxygen atom in the chymostatin/SGPA complex is not associated with newly bound solvent molecules. As a consequence of the conformation taken by the aldehyde group of chymostatin, a close contact with the imidazole ring of His-57 does not develop. Secondly, due to close contacts which would develop

with the terminal methyl groups of the P2 leucyl side chain, movement of the imidazole ring of His-57 from its native position would not be possible when chymostatin is bound.

In summary, the two inhibitor/SGPA complexes examined indicate that formation of a covalent tetrahedral hemiacetal adduct with Ser-195 is a consistent feature of specific peptide aldehyde complexation with SGPA. Further, the overall binding mode of residues removed from the active aldehyde function is quite similar. However, the resulting conformation of His-57 in such complexes appears to be dependent on the type of residues that make up the polypeptide chain backbone of the specific peptide aldehyde inhibitor bound.

VIII. Chloromethyl Ketone Inhibitor Studies Of SGPB

A. Chloromethyl Ketone Peptide Analogs

Chloromethyl ketone peptide analogs of good substrates have been shown to be covalently bound, irreversible inhibitors of serine proteases (Shaw, 1970; Powers, 1977). The usefulness of such inhibitors was first demonstrated by Schoellmann and Shaw (1963) in an experiment involving TPCK inhibition of alpha-chymotrypsin. These peptide analogs have a C-terminal chloromethyl ketone group rather than the terminal carboxyl group of normal peptides. The first step in the reaction of serine proteases with substrate analogs of this type, apparently involves the formation of an initial enzyme/inhibitor complex in which the inhibitor is recognized by the specific interactions of its P1 residue in the primary specificity site (Powers, 1977). Indeed, chloromethyl ketone peptides which do not have a P1 residue complementary to the primary specificity binding site are poor inhibitors. Following initial complexation, irreversible inhibition occurs upon the formation of a covalent bond between the active site histidine residue and the methylene group of the inhibitor.

Both chemical and structural evidence have suggested that the formation of a tetrahedral hemiketal with the catalytic serine residue of serine proteases is a prerequisite in the alkylation of the active site histidine residue (Poulos et al., 1976). For example, conversion of

Ser-195 in alpha-chymotrypsin to a dehydroalanine residue (Weiner et al., 1966) renders this enzyme unreactive towards chloromethyl ketone inhibitors. Crystallographic analysis of subtilisin - chloromethyl ketone inhibitor complexes have also shown that these inhibitors form not only a covalent bond to the reactive histidine in the active site, but a further covalent bond to the catalytic serine residue is present. These results have been taken to suggest that only upon hemiketal formation is the chloromethyl ketone moiety suitably positioned with respect to the reactive histidine residue in order for the alkylation process to proceed (Poulos et al., 1976).

The structural aspects of chloromethyl ketone inhibition of gamma-chymotrypsin (Segal et al., 1971) and of subtilisin (Robertus, et al., 1972a; Poulos, et al., 1976) have been investigated. In both cases, these studies were able to demonstrate the overall mode of inhibitor binding and to reveal the positions of binding subsites in the active site. In each enzyme/inhibitor complex, the polypeptide backbone of the inhibitor is bound in an anti-parallel beta sheet fashion to the enzyme polypeptide strand which forms the external surface of the primary specificity pocket. This binding mode has been taken as being representative of true substrate binding and agrees well with that observed for other kinds of inhibitors (Ruhlmann et al., 1973; Sweet et al.,1974).

Studies of chloromethyl ketone peptide inhibition of

SGPB have also been carried out using a number of inhibitors differing in overall length and amino acid composition (Gertler, 1974). It has been shown that such inhibitors bind covalently to the active site histidine residue of SGPB. This study was also able to partially map out the specificity and size of the substrate binding region of this enzyme. To further determine the nature of chloromethyl ketone inhibition of SGPB, crystallographic analyses of two SGPB-inhibitor complexes have been carried out to 2.8 angstrom resolution. The two chloromethyl ketone peptides used were: Boc-Ala-Gly-Phe-CK and Boc-Gly-Leu-Phe-CK. These are referred to in the following text as the AGF and GLF inhibitors, respectively. Inhibited samples of SGPB were generously supplied for these experiments by Dr. A. Gertler.

B. Crystallization and Data Collection

Like crystals of native SGPB, suitable crystals of each SGPB - chloromethyl ketone inhibited complex were obtained from 0.7M KH(2)PO(4) at pH 4.2. Precession camera photography established that complex crystals retained the same space group (P2(1)2(1)2) and were isomorphous with the crystal form used to solve the structure of native SGPB. Diffractometer data to 2.8 angstrom resolution for each SGPB/inhibitor complex were collected, processed and scaled following the methodology described earlier for SGPA/inhibitor complexes. Relevant crystallographic data collection and processing statistics for native SGPB and the

TABLE 28

SGPB: Native And Inhibitor Complex Diffraction Statistics

Data	Native	GLF	AGF
a	44.16(2)	44.22(2)	44.19(1)
b	108.91(5)	108.72(5)	108.54(4)
c	37.34(2)	37.31(2)	37.28(1)
No. of reflections measured	8759	4992	4978
Max. absorption correction (%)	26.7	50.4	20.5
Max. crystal decay (%)	6.2	1.5	3.7
Percent reflections (I > 3sigma(I))	92.2	88.4	77.5
Absolute scale	13.07	11.57	22.12
Overall isotropic B (angstroms)^2	11.8	12.7	12.7
^{1}R(I) (%)	-	14.8	16.9

^{1}R(I) is calculated in the same manner as R(D) (Table 7), using inhibitor complex rather than heavy-atom derivative amplitudes.

two SGPB - chloromethyl ketone complexes examined, are given in Table 28.

Calculated structure factor amplitude differences were used to compute a difference Fourier electron density map of each SGPB/inhibitor complex. Coefficients for each difference map, were the figure of merit weighted differences, {F(P+I) - F(P)} exp(i alpha(P)), where F(P+I) represents the structure factor amplitudes obtained from the crystal of the inhibited complex of SGPB and F(P) represents the structure factor amplitudes of the native SGPB crystal. The native SGPB diffraction data used in this study was that

of Delbaere et al. (1975). The phases and figures of merit for each difference map were also obtained from the 2.8 angstrom resolution multiple isomorphous replacement phase determination of native SGPB.

Each SGPB-inhibitor difference map was interpreted in a Richards optical comparator (Richards, 1968). After an optimal fit of the Watson-Kendrew model of each inhibitor into the difference electron density had been achieved, coordinates for all non-hydrogen inhibitor atoms were measured using the plumb-line method. These coordinates were subsequently refined using Diamond's (1966, 1974) model building procedure (GLF: r.m.s.=0.30 angstroms; AGF: r.m.s.=0.22 angstroms). Final fits of the model built coordinates were made using the MMS-X graphics system.

Also used in providing an additional check on the interpretation of each difference electron density map, were Fourier maps using as coefficients {2F(P+I) - F(P)} and the figures of merit and phases of native SGPB. These maps, which approximate the overall electron density distribution of the inhibitor/SGPB complexes, were used in both the optical comparator and on the MMS-X graphics system to further guide the interpretation of the difference electron density maps.

C. Interpretation of Difference Electron Density Maps

Both of the SGPB - chloromethyl ketone complexes have difference electron density maps that are similar in overall appearance. The main feature of each difference map is a well defined continuous chain of positive electron density lying in the substrate binding cleft. This chain originates in the vicinity of the catalytic residues His-57 and Ser-195. Figure 49 shows the final fitted models of the GLF and AGF inhibitors in their respective difference electron densities. For clarity, only the positive contour envelope is shown in this Figure. Also shown in this Figure are the side chains of the active site residues His-57 and Ser-195. Both difference electron density maps were easily interpreted in terms of the known chemical structure of these inhibitors. The absence of significant peaks of difference electron density other than those associated with the bound inhibitors, indicates that SGPB does not undergo large conformational changes upon inhibitor binding. Indeed, only very small movements in a few enzyme atomic positions are indicated and these are restricted to residues interacting directly with the bound inhibitors.

Comparison of Figures 49a and 49b illustrates the different amino acid compositions of the two inhibitors studied. Particularly prominent is the presence of difference electron density at the P2 leucine residue of the GLF inhibitor and the absence of any side chain density at the equivalent glycine residue of the AGF inhibitor. Also

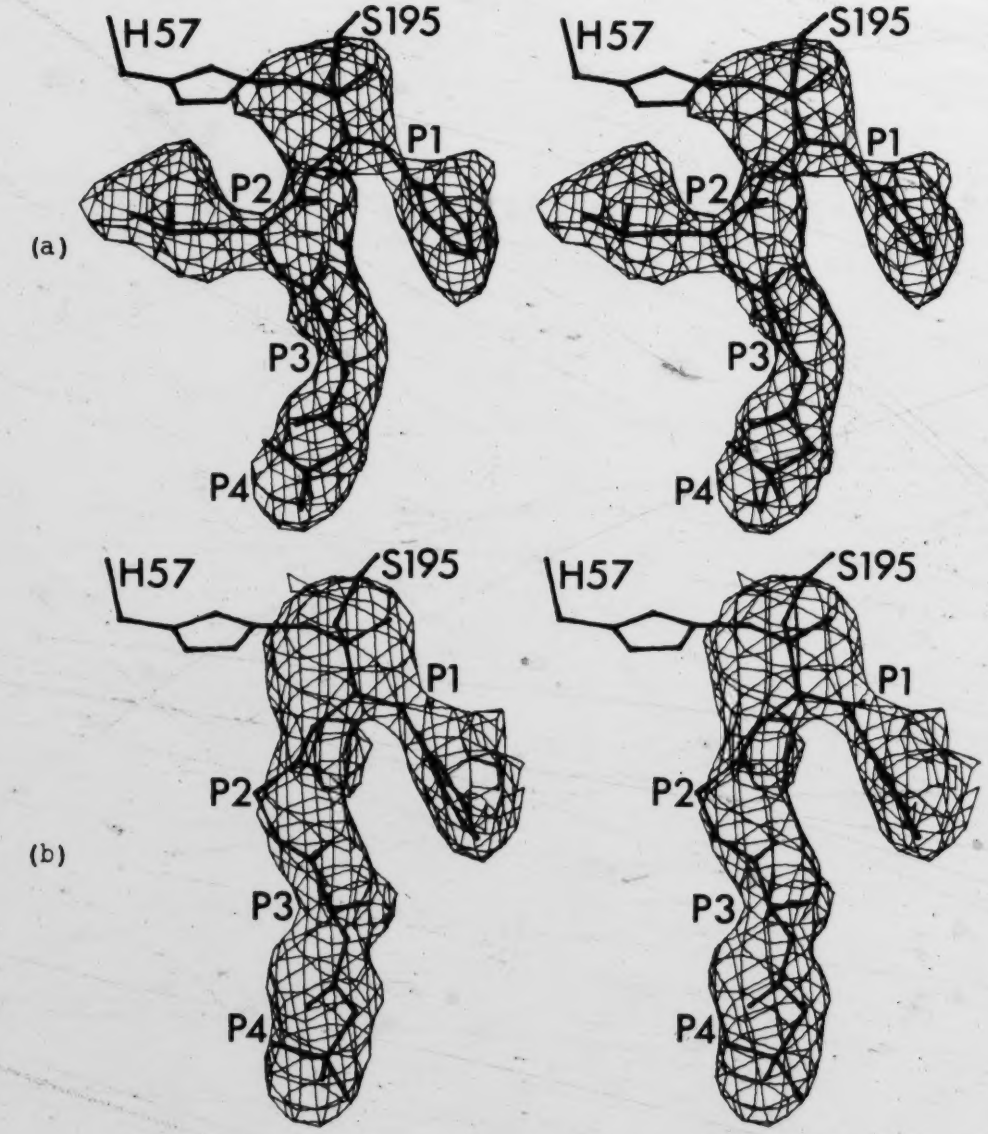


Fig. 49. Stereo-representation of the 2.8 angstrom resolution difference electron density maps of (a) the GLF/SGPB and (b) the AGF/SGPB inhibitor complexes. Superimposed in their respective difference electron density, are the final fitted molecular models of the GLF and AGF inhibitors. The side chains of His-57 and Ser-195 are also drawn. The positive contour envelope shown for each inhibitor is drawn at a level of 0.078e/(angstroms)^3. The standard error of these maps have been estimated to be (a) 0.043e/(angstroms)^3 for the GLF inhibitor and (b) 0.056e/(angstroms)^3 for the AGF inhibitor.

clearly visible, is the exchange of the P3 glycine residue of the GLF inhibitor for a P3 alanine residue in the AGF inhibitor. In both inhibitor complexes the polypeptide backbone of each inhibitor is bound to the surface of SGPB in an anti-parallel beta sheet conformation involving three hydrogen bonds. Although the polypeptide backbones of both inhibitors are bound in very similar conformations, small differences are observed. These likely arise from the different amino acid compositions of the two inhibitors.

The GLF/SGPB Complex

Figure 50 shows the conformation of the GLF chloromethyl ketone inhibitor bound in the active site of SGPB, as interpreted from the difference electron density of Figure 49a. The difference electron density of this inhibitor overlaps with the native SGPB electron density at only two positions: the NE2 nitrogen atom of the side chain of His-57 and at the OG oxygen atom of the side chain of Ser-195. Interpretation of the difference electron density map in this region, indicates two covalent linkages have been formed from the inhibitor to these two enzyme atoms. The NE2 nitrogen atom of His-57 is found covalently linked to the C-terminal methylene carbon atom of the inhibitor. A second covalent bond is formed from the carbonyl carbon atom of the P1 phenylalanine residue to the OG oxygen atom of Ser-195. Formation of this latter bond causes the carbonyl carbon atom to take a tetrahedral geometry. No movement of

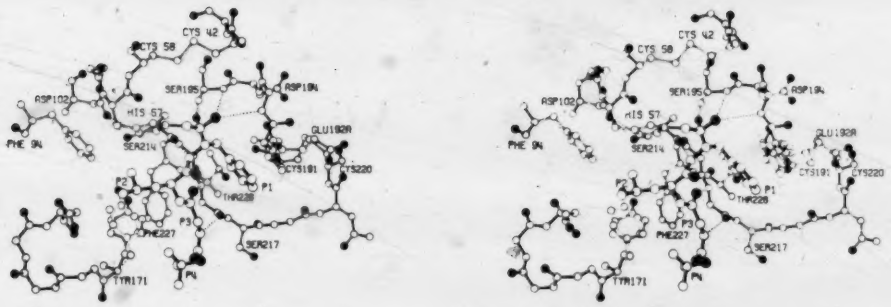


Fig. 50. Stereo-drawing of the GLF inhibitor bound in the active site of SGPB, as interpreted from the 2.8 angstrom resolution difference electron density map of the GLF/SGPB complex. Polypeptide main chain bonding of the enzyme and all interatomic bonds of the bound inhibitor are indicated by filled black bonds. All oxygen atoms are distinguished by filled black circles. Hydrogen bonds to active site residues and the bound inhibitor are indicated by thin dashed lines. The position of the side chain of Tyr-171 in this complex is also drawn. The original native conformation of the side chain of this residue is shown by dashed lines.

either His-57 or Ser-195 from their native positions is indicated in the difference electron density map.

Several lines of evidence support the interpretation of the presence of two covalent bonds being formed between the GLF inhibitor and active site residues of SGPB. Three major observations support the premise for a covalent bond from the gamma oxygen of Ser-195 to the carbonyl carbon atom of the P1 phenylalanine residue of the inhibitor. Firstly, if the difference electron density is fit by a planar carbonyl carbon, as if a covalent bond did not exist, the fit to the difference electron density map is poor. Also, the resultant planar carbonyl carbon atom position remains in very close

proximity to the gamma oxygen atom of Ser-195 (approximately 1.7 angstroms). Secondly, the difference electron density of the bound inhibitor overlaps with the native electron density present for the side chain of Ser-195 indicating that a bond has been formed (Figure 49a). In addition, the positioning of a tetrahedral sp^3 carbon atom instead of a planar sp^2 carbonyl carbon atom in the difference electron density greatly improves the overall fit of the inhibitor. Thirdly, a figure of merit weighted {2F(P+I) - F(P)} exp(i alpha(P)) electron density map has been calculated for this inhibitor complex. A portion of this electron density map encompassing the active site residues of SGPB is shown in Figure 51. The presence of continuous electron density between the inhibitor P1 carbonyl carbon atom and the gamma oxygen atom of Ser-195 in this map, confirms the existence of a bonding interaction between these atoms.

Covalent bond formation between the methylene carbon atom of the GLF inhibitor and the His-57 NE2 nitrogen atom is also supported in Figures 49a and 51. In Figure 49a, difference electron density representing the bound conformation of the inhibitor overlaps with that of the imidazole ring of His-57. Further evidence for covalent bond formation is present in Figure 51, where it is shown that there is continuous electron density between the methylene carbon atom position of the inhibitor and the NE2 nitrogen atom of His-57.

As shown in Figure 50, the carbonyl oxygen atom of P1

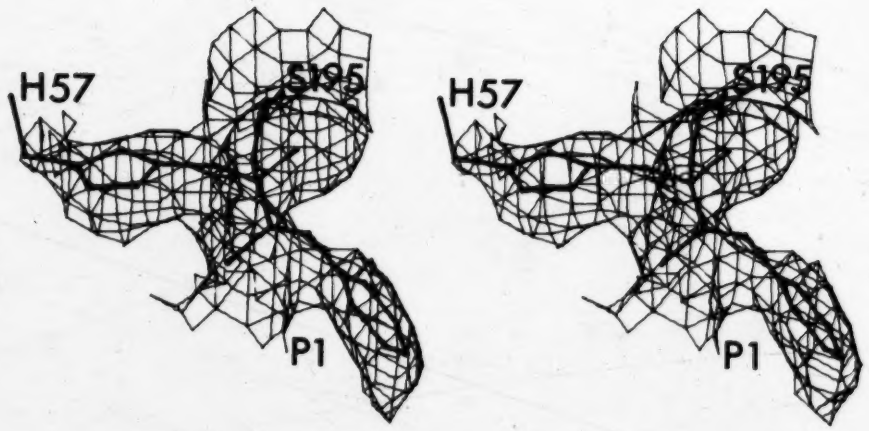


Fig. 51. Stereo-representation of the electron density of the GLF/SGPB complex in the vicinity of the active site residues, His-57 and Ser-195. The P1 residue of the bound inhibitor as well as the side chains of Ser-195 and His-57 are also drawn. The contour envelope presented is drawn at 0.35e/(angstroms)^3.

phenylalanine is oriented into the oxyanion hole, where it forms two hydrogen bonds, one to each of the amide groups of Ser-195 and Gly-193. The side chain of P1 phenylalanine lies in the primary specificity binding site, which in SGPB is a shallow surface pocket near the catalytic center. This binding subsite (S1) is defined by the polypeptide backbone of residues 192 to 192B and residues 215 to 218, as well as the plane of the proline ring of residue 192B and at the bottom by Thr-226. S1 subsite interactions are completed by the formation of a hydrogen bond from the amide group of P1 phenylalanine to the carbonyl oxygen atom of Ser-214.

The side chain of P2 leucine lies in a well defined binding subsite (S2) to the other side of the substrate binding cleft (Figure 50). This subsite is formed from the

side chains of His-57, Phe-94 and Tyr-171. Portions of the polypeptide chain of residues Tyr-171 to Val-176 also form one side of this subsite. The side chain of Tyr-171 is found rotated in the GLF inhibitor complex in order to relieve close contacts that would otherwise be formed with the P2 leucyl side chain. The new side chain conformation of Tyr-171 in the SGPB/GLF inhibitor complex is shown in Figure 50.

Two hydrogen bonds are formed by P3 glycine to Gly-216 in the S3 binding subsite. The N-terminal Boc group (P4) of the GLF inhibitor is found at the extremity of the substrate binding region in a surface hydrophobic depression. Enzyme side chains in this region include Val-169 and Phe-227. It seems likely that further residues of longer inhibitors or substrates would lie off the surface of SGPB and protrude into the surrounding solvent.

The AGF/SGPB Complex

The chloromethyl ketone inhibitor AGF as bound in the active site of SGPB is shown in Figure 52. Based on the same criteria as discussed previously, this inhibitor is also covalently bound to the active site residues, Ser-195 and His-57, of SGPB. The {2F(P+I) - F(P)} electron density map of the SGPB/AGF complex is shown in Figure 53. As comparison of Figures 50 and 52 shows, the AGF and GLF inhibitors both form very similar anti-parallel beta sheet conformations in the active site region.

The replacement of P2 leucine in the GLF inhibitor by a

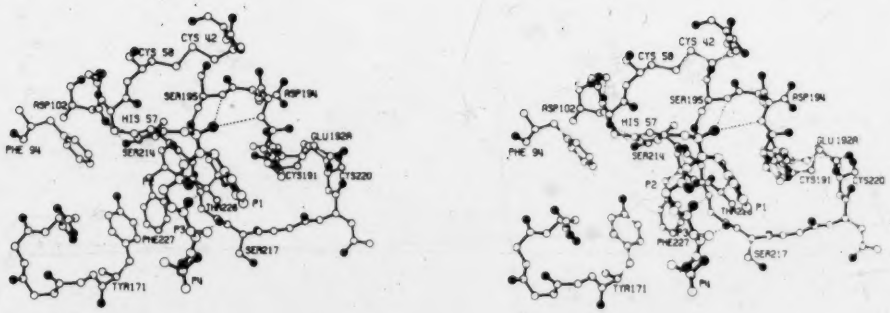


Fig. 52. Stereo-drawing of the active site of SGPB showing the conformation of the bound AGF inhibitor. This inhibitor was positioned by interpretation of the 2.8 angstrom resolution difference electron density map of the AGF/SGPB complex (Figure 49b). This drawing is presented with the same view as Figure 50.

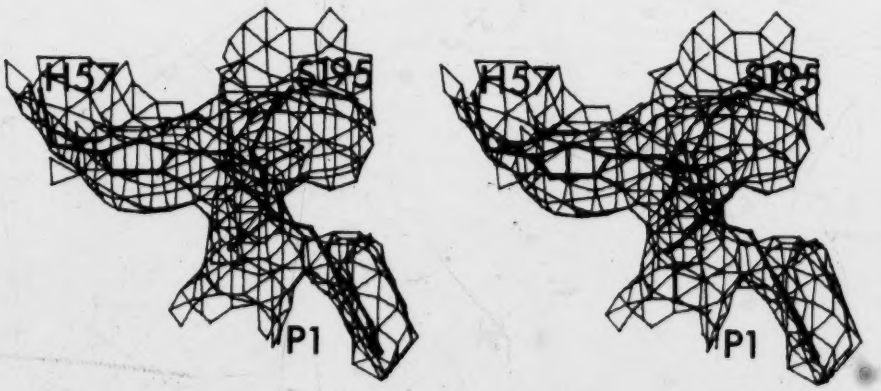


Fig. 53. Stereo-representation of the electron density map of the AGF/SGPB complex in the active site region. This map was calculated and is oriented in a similar manner as that shown in Figure 51. The superimposed model represents the side chains of His-57 and Ser-195, as well as the P1 residue of the bound inhibitor.

glycine residue in the AGF inhibitor is the most significant structural difference between these inhibitors. Comparisons of the bound conformations of the two inhibitors about the

S2 binding subsite, show them to bind in a very similar manner. However, in the AGF complex, the absence of a P2 leucyl side chain results in the retention of the native conformation of the side chain of Tyr-171. The only other chemical difference between the two bound chloromethyl ketones is the exchange of a P3 glycine residue in the GLF inhibitor, for an alanine residue in the AGF inhibitor. As Figure 52 shows, this alanyl side chain is oriented away from the enzyme surface and there is no apparent binding pocket for side chains in the S3 subsite.

Appendix 4 contains a list of coordinates for the GLF inhibitor non-hydrogen atoms and the two active site residues to which the inhibitor is covalently attached. Also included, are the reoriented Tyr-171 side chain coordinates. Appendix 5 contains a list of coordinates for the non-hydrogen atoms of the AGF inhibitor, as well as those of Ser-195 and His-57, to which it is covalently bound. Coordinates in both Appendices 4 and 5 are in orthogonal angstrom units corresponding to the crystallographic unit cell of SGPB.

D. Chloromethyl Ketone Peptides As Substrate Analogs

Covalent bond formation between the active site histidine residue of serine proteases and the methylene carbon atom of chloromethyl ketone inhibitors has been well characterized in a number of chemical (Shaw, 1970; Powers, 1977) and structural studies (Robertus et al., 1972a; Segal

et al., 1971; Poulos et al., 1976). Inhibition studies of chloromethyl ketone inhibitor complexes with SGPB (Gertler, 1974) have also shown that the active site histidine residue of this enzyme is bound covalently to such inhibitors. The present study has examined the three-dimensional conformation of two chloromethyl ketone inhibitors bound on the surface of SGPB. In each case, covalent bond formation to His-57 of this enzyme was demonstrated. Thus, this aspect of inhibitor complexation to SGPB agrees with earlier studies, and is consistent with results observed for inhibitor binding to other serine proteases.

Only recently, has a covalent bond between the carbonyl carbon atom of the P1 residue of a bound chloromethyl ketone peptide inhibitor and the gamma oxygen atom of the catalytic serine residue of serine proteases been proposed (Poulos et al., 1976). Previous non-structural studies related to determining the point of attachment of these inhibitors did not detect the existence of this bond. Similar experiments conducted with SGPB/inhibitor complexes (Gertler, 1974), were also unable to detect bond formation to Ser-195. Nevertheless, in agreement with the results of Poulos et al. (1976), the present structural study has established the presence of such a bond in SGPB/inhibitor complexes.

This discrepancy between earlier inhibition studies and more recent crystallographic analyses of chloromethyl ketone complexes, apparently arises since bond formation to Ser-195 can only be observed when the unique structures present in

the active sites of serine proteases are preserved intact. Thus, the hemiketal complex formed is dependent upon the presence of stabilizing forces arising from surface interactions in the active site. Some of these interactions are likely to include: the juxtaposition of nearby active site residues; interactions in the oxyanion hole; the primary specificity pocket and interactions formed by residues further removed from the active site region. Destruction of the unique conformation of the active site region, such as that occurring in non-structural methods employed to detect the presence of a covalent bond to the catalytic histidine residue, results in the dissolution of the hemiketal complex before it can be detected (Schoellmann and Shaw, 1963; Gertler, 1974).

A number of studies have pointed out the importance of specific chloromethyl ketone inhibitor interactions in the inhibition of serine proteases (Shaw, 1970; Powers, 1977). In this regard SGPB is similar. Gertler (1974) has shown that only specific chloromethyl ketone peptides are effective inhibitors of SGPB. Also, the length and amino acid composition of such inhibitor peptides can have pronounced effects on their ability to inhibit SGPB. In light of the specific nature of inhibition and the presence of hemiketal formation, Poulos *et al.* (1976) have proposed that chloromethyl ketone studies mimic the structure of true tetrahedral substrate complexes. The observation of hemiketal structures in SGPB complexes and their similarity

to postulated tetrahedral substrate intermediates (Figure 2) supports the proposals of Poulos et al. (1976). However, close parallels between the observed inhibitor complexes and the transient species of true substrate complexes should not be liberally drawn. The simple presence of a covalent bond from these inhibitors to the catalytic histidine residue is likely to cause reorientation of the hemiketal group from the tetrahedral structure present during substrate catalysis.

The specific nature of chloromethyl ketone complexation has also suggested that the peptide portions of these inhibitors are bound in surface binding subsites in a manner similar to true substrates. Indeed, crystallographic studies of inhibitor complexes of gamma-chymotrypsin and subtilisin have lead to reasonable models of substrate binding for these enzymes (Segal et al, 1971; Robertus et al., 1972a). A number of aspects of the present binding study also serve to explain observations resulting from solution studies of SGPB. For example, both Narahashi (1972) and Bauer (1978) have found the primary specificity pocket of SGPB preferentially binds the side chains of phenylalanine, tyrosine and leucine. In each of the GLF and AGF inhibitor complexes with SGPB (Figures 50 and 52) a phenylalanyl side chain is bound in the primary specificity pocket. In agreement with solution studies, the overall size and shape of this subsite is well suited for the side chain of phenylalanine. The non-planar leucyl or the larger tyrosyl

side chains could also be accommodated. However, model building studies using the bound conformation of the GLF inhibitor as a guide, indicate that the side chain of tryptophan is too large to be bound in this binding pocket. This is in good agreement with solution studies of substrates containing this residue at the P1 position (Bauer, 1978).

Other studies have also indicated that SGPB has an extended substrate binding region. The importance of these secondary binding sites is evident from the inability of SGPB to cleave very short substrates at appreciable rates (Bauer, 1978). Also, slow rates of inhibition are observed with chloromethyl ketone peptides of small size (Gertler, 1974). The present study can identify the three secondary binding subsites (S2-S4) beyond the primary specificity site. Reference to Figures 50 and 52 shows that substantial contacts are made in these binding subsites. Indeed, subsite S2 appears to be nearly as well defined as the primary specificity pocket. The significance of contacts formed in the S2 subsite is shown by the approximately 80 fold increase in the rate of inhibition upon using the chloromethyl ketone inhibitor Ac-Leu-Phe-CK as opposed to Ac-Ala-Phe-CK (Gertler, 1974). A similar increase in inhibition rate (approximately 100 fold) is observed between the two inhibitors examined in this study. As illustrated in Figure 50, the leucyl side chain of the GLF inhibitor forms many contacts in the S2 subsite.

The S3 binding subsite of SGPB is without a side chain binding pocket, but does form two hydrogen bonds to the bound polypeptide backbone of each inhibitor. Bauer (1978) has shown that this subsite plays an important role in orienting substrates in the active site. In the absence of a side chain binding pocket, it appears that this enhancement arises chiefly from the hydrogen bonded contacts formed in this subsite. Gertler (1974) was able to demonstrate that the S4 subsite is also a determinant in orienting chloromethyl ketone inhibitors by showing that a P4 Boc group results in greater inhibitor affinity than a smaller P4 N-acetyl group. Substrate studies also indicate the S4 subsite plays a role in binding and orienting substrate peptides (Bauer, 1978). Examination of the active site region of SGPB (Figures 50 or 52) indicates that the S4 binding subsite is quite remote from the active site residues of this enzyme. In agreement with this observation, Bauer (1978) has noted that there are probably no substantial enzyme - substrate interactions beyond the S4 subsite.

Comparison of Figures 50 and 52 with Figures 43 and 47, shows that the close structural homology between SGPA and SGPB extends to the manner in which each enzyme binds inhibitors on their respective surfaces. Like SGPA, the overall binding mode of peptide inhibitors is in an anti-parallel beta sheet conformation. This is a characteristic shared with the pancreatic counterpart of

these enzymes (Segal et al., 1971). As observed for SGPA, the primary specificity pocket of SGPB is less of a well defined binding subsite than that of alpha-chymotrypsin, explaining the specificity differences observed between the microbial and pancreatic enzymes. The presence of well defined secondary binding subsites in SGPA and SGPB, also explains the greater length dependence that these enzymes show relative to alpha-chymotrypsin. It has been suggested (Bauer, 1978) that such secondary subsite interactions are necessary to compensate for the less specific nature of the primary specificity pockets of these enzymes. The results of the present structural studies support this proposal in terms of the observed structural attributes of the binding subsites of these enzymes.

In conclusion, the binding modes observed for the chloromethyl ketone peptides bound to SGPB in this study serve to explain observations of earlier substrate and inhibitor solution studies. On this basis, it seems likely that the bound conformation of these inhibitors is representative of that expected for true substrates. Further, the present study confirms the active site conformation of chloromethyl ketone inhibition as observed in the earlier study of Poulos et al. (1976). Several conclusions can also be drawn in comparisons with structural studies of inhibitor complexes of other serine proteases. Firstly, as pointed out in earlier solution studies (Bauer, 1978), the mode of inhibitor binding and the structures of

the binding subsites of both SGPA and SGPB are very similar. Secondly, both microbial enzymes bind inhibitors in a manner similar to gamma-chymotrypsin (Segal et al., 1971), alpha-chymotrypsin (Sweet et al., 1973) and trypsin (Ruhlmann et al., 1974). Finally, structural studies of subtilisin/inhibitor complexes (Robertus et al., 1972a; Poulos et al., 1976) have shown that even the structurally unrelated subtilisins invoke a peptide binding mode remarkably similar to that observed for the pancreatic and microbial pancreatic-like serine proteases.

References

Adams, M.J., Haas, D.J., Jeffery, B.A., McPherson, A. Jr., Mermall, H.L., Rossmann, M.G., Schevitz, R.W. and Wonacott, A.J. (1969). J. Mol. Biol. 41, 159-188.

Aoyagi, T. and Umezawa, H. (1975). 'In' Proteases and Biological Control (Reich, E., Rifkin, D.B. and Shaw, E., eds), pp. 429-454, Cold Spring Harbor Laboratory, New York.

Awad, W.M., Soto, A., Siegal, S., Skiba, W., Bernstrom, G. and Ochoa, M. (1972). J. Biol. Chem. 247, 4144-4154.

Bachovchin, W.W. and Roberts, J.D. (1978). J. Amer. Chem. Soc. 100, 8041-8047.

Bauer, C.-A. (1976). Biochem. Biophys. Acta 438, 495-502.

Bauer, C.-A. (1977). Acta Chem. Scand. B31, 637-639.

Bauer, C.-A. (1978). Biochemistry 17, 375-380.

Bauer, C.-A. and Lofqvist, B. (1973). Acta Chem. Scand. 27, 3147-3166.

Bauer, C.-A. and Pettersson, G. (1974). Eur. J. Biochem. 45, 469-472.

Bauer, C.-A., Lofqvist, B. and Pettersson, G. (1974). Eur. J. Biochem. 41, 45-49.

Bauer, C.-A., Thompson, R.C. and Blout, E.R. (1976a). Biochemistry 15, 1291-1295.

Bauer, C.-A., Thompson, R.C. and Blout, E.R. (1976b). Biochemistry 15, 1296-1299.

Bender, M.L. and Kezdy, F.J. (1964). J. Amer. Chem. Soc. 86, 3704-3714.

Bender, M.L. and Killheffer, J.V. (1973). CRC Crit. Rev. Biochem. 1, 149-199.

Biochem. J. (1969). 113, 1-4.

Birktoft, J.J. and Blow, D.M. (1972). J. Mol. Biol. 68, 187-240.

Birktoft, J.J., Kraut, J. and Freer, S.T. (1976). Biochemistry 15, 4481-4485.

Blake, C.C.F., Fenn, R.H., Koenig, D.F., North, A.C.T., Phillips, D.C. and Poljak, R.J. (1963). Acta Crystallogr. 16, A77.

Blow, D.M. (1976). Acc. Chem. Res. 9, 145-152.

Blow, D.M. and Crick, F.H.C. (1959). Acta Crystallogr. 12, 794-802.

Blundell, T.L. and Johnson, L.N. (1976). 'In' Protein Crystallography, pp 416-417, Academic Press, New York.

Bode, W. and Schwager, P. (1975). J. Mol. Biol. 98, 693-717.

Bode, W., Schwager, P. and Huber, R. (1976). 'In' Proteolysis and Physiological Regulation (Miami Winter Symposia) (Ribbons, D.W. and Brew, K. eds), vol. 11, pp. 43-76, Academic Press, New York.

Breaux, E.J. and Bender, M.L. (1975). FEBS Lett. 56, 81-84.

Buchanan, R.E. and Gibbons, N.E. (1974). 'In' Bergeys Manuel of Determinative Bacteriology , eighth edition, pp. 747-750, Williams and Wilkins Company, Baltimore.

Chen, R., Gorenstein, D.G., Kennedy, W.P., Lowe, G., Nurse, D. and Schultz, R.M. (1979). Biochemistry 18, 921-926.

Christensen, P. and Cook, F.D. (1978). Int. J. Sys.

Bacteriol. 28, 367-393.

Clark, P.I., Lowe, G. and Nurse, D. (1977). J. Chem. Soc. Chem. Commun., 451-453.

Codding, P.W., Delbaere, L.T.J., Hayakawa, K., Hutcheon, W.L.B., James, M.N.G. and Jurasek, L. (1974). Can. J. Biochem. 52, 208-220.

Cotton, F.A. and Wilkinson, G. (1972). 'In' Advances in Inorganic Chemistry, 3rd edition, pp. 972-990, Interscience, New York.

Cruickshank, D.W.J. (1949). Acta Crystallogr. 2, 65-82.

Cullis, A.F., Muirhead, H., Perutz, M.F., Rossmann, M.G. and North, A.C.T. (1961). Proc. Roy. Soc. ser. A, 265, 15-38.

Davies, D.R., Cohen, G.H., Silverton, E.W., Braxton, H.P. and Matthews, B.W. (1969). Acta Crystallogr. A25, S-182.

Deisenhofer, J. and Steigemann, W. (1975). Acta Crystallogr. B31, 238-250.

Delbaere, L.T.J., Hutcheon, W.L.B., James, M.N.G., and Thiessen, W.E. (1975). Nature 257, 758-763.

Diamond, R. (1966). Acta Crystallogr. 21, 253-266.

Diamond, R. (1974). J. Mol. Biol. 82, 371-391.

Dickerson, R.E., Kendrew, J.C. and Strandberg, B.E. (1961). Acta Crystallogr. 14, 1188-1195.

Dickerson, R.E., Kopka, M.L., Varnum, J.C. and Weinzierl, J.E. (1967). Acta Crystallogr. 23, 511-522.

Dickerson, R.E., Weinzierl, J.E. and Palmer, R.A. (1968). Acta Crystallogr. B24, 997-1003.

Drenth, J., Hol, W.G.J., Jansonius, J.N. and Koekoek, R. (1972). Eur. J. Biochem. 26, 177-181.

Egan, W., Shindo, H. and Cohen, J. (1977). Ann. Rev. Biophys. Bioeng. 6, 383-417.

Fehlhammer, H. and Bode, W. (1975). J. Mol. Biol. 98, 683-692.

Fehlhammer, H., Bode., W. and Huber, R. (1977). J. Mol. Biol. 111, 415-438.

Feinstein, G., Malemud, C.J. and Janoff, A. (1976). Biochim. Biophys. Acta 429, 925-932.

Fersht, A. (1977). Enzyme Structure and Mechanism, pp. 175-186, Freeman and Company, San Francisco.

Freer, S.T., Kraut, J., Robertus, J.D., Wright, H.T. and Xuong, Ng.-H. (1970). Biochemistry 9, 1997-2009.

Gertler, A. (1974). FEBS Lett. 43, 81-85.

Gertler, A. and Trop, M. (1971). Eur. J. Biochem. 19, 90-96.

Gillespie, D.C. and Cook, F.D. (1965). Can. J. Microbiol. 11, 109-118.

Gorenstein, D.G., Kar, D. and Momii, R.K. (1976). Biochem. Biophys. Res. Commun. 73, 105-111.

Hartley, B.S. and Kauffman, D.L. (1966). Biochem. J. 101, 229-231.

Hartley, B.S. and Shotton, D.M. (1971). 'In' The Enzymes (Boyer, P.D., ed.), vol. 3, third edition, pp. 323-373, Academic Press, New York.

Hartley, B.S., Burleigh, B.D., Midwinter, G.G., Moore, C.H., Morris, H.R., Rigby, P.W.J., Smith, M.J. and Taylor,

S.S. (1972). 'In' Enzymes: Structure and Function, FEBS 8th Meet., Amsterdam (Drenth, J., Oosterbaan, R.H. and Veeger, C. eds), pp. 151-176, North-Holland, Amsterdam.

Henderson, R. (1970). J. Mol. Biol. 54, 341-354.

Henderson, R. and Moffat, J.K. (1971). Acta Crystallogr. B27, 1414-1420.

Hendrickson, W.A., Love, W.E. and Karle, J. (1973). J. Mol. Biol. 74, 331-361.

Hess, G.P. (1971). 'In' The Enzymes (Boyer, P.D., ed.), vol. 3, third edition, pp. 213-248, Academic Press, New York.

Hill, E.J. and Banaszak, L.J. (1973). Acta Crystallogr. B29, 372.

Hiramatsu, A. and Ouchi, T. (1963). J. Biochem. (Tokyo) 54, 462-464.

Huber, R., Kukla, D., Bode, W., Schwager, P., Bartels, K., Deisenhofer, J. and Steigemann, W. (1974). J. Mol. Biol. 89, 73-101.

Hunkapiller, M.W., Smallcombe, S.H., Whitaker, D.R. and Richards, J.H. (1973). Biochemistry 12, 4732-4743.

Hunkapiller, M.W., Smallcombe, S.H. and Richards, J.H. (1975). Org. Mag. Reson. 7, 262-265.

Ito, A., Tokawa, K. and Shimizu, B. (1972). Biochem. Biophys. Res. Commun. 49, 343-349.

James, M.N.G. and Smillie, L.B. (1969). Nature 224, 694-695.

Jansen, E.F., Nutting, M.D.F., Jang, R. and Balls, A.K. (1949). J. Biol. Chem. 179, 189-199.

Johnson, P. and Smillie, L.B. (1971). Can. J. Biochem. 49,

548-562.

Johnson, P. and Smillie, L.B. (1974). FEBS Lett. 47, 1-6.

Jurasek, L., Carpenter, M.R., Smillie, L.B., Gertler, A., Levy, S. and Ericsson, L.H. (1974). Biochem. Biophys. Res. Commun. 61, 1095-1100.

Jurasek, L., Johnson, P., Olafson, R.W. and Smillie, L.B. (1971). Can. J. Biochem. 49, 1195-1201.

Kaplan, H. and Dugas, H. (1969). Biochem. Biophys. Res. Commun. 34, 681-685.

Kaplan, H. and Whitaker, D.R. (1967). J. Amer. Chem. Soc. 89, 3352-3353.

Kaplan, H. and Whitaker, D.R. (1969). Can. J. Biochem. 47, 305-316.

Kaplan, H., Symonds, V.B., Dugas, H. and Whitaker, D.R. (1970). Can. J. Biochem. 48, 649-658.

Katznelson, H. and Henderson, V.E. (1962). Can. J. Microbiol. 8, 875-882.

Katznelson, H. and Henderson, V.E. (1964). Can. J. Microbiol. 10, 37-41.

Katznelson, H., Gillespie, D.C. and Cook, F.D. (1964). Can. J. Microbiol. 10, 699-704.

Keil, B. (1971). 'In' The Enzymes (Boyer, P.D., ed.), vol. 3, third edition, pp. 249-275, Academic Press, New York.

Kossiakoff, A., Chambers, J., Kay, L. and Stroud, R. (1977). Biochemistry 16, 654-664.

Kraut, J. (1971). 'In' The Enzymes (Boyer, P.D., ed.), vol. 3, third edition, pp. 547-560, Academic Press, New York.

Kraut, J. (1977). Ann. Rev. Biochem. 46, 331-358.

Kraut, J., Robertus, J.D., Birktoft, J.J., Alden, R.A., Wilcox, P.E. and Powers, J.C. (1971). Cold Spring Harbor Symp. Quant. Biol. 36, 117-123.

Kraut, J., Sieker, L.C., High, D.F. and Freer, S.T. (1962). Proc. Natl. Acad. Sci. (USA) 48, 1417-1424.

Krieger, M., Chambers, J.L., Christoph, G.G., Stroud, R.M. and Trus, B.L. (1974). Acta Crystallogr. A30, 740-748.

Lenhert, P.G. (1975). J. Appl. Crystallogr. 8, 568-570.

Lewis, C.A. Jr. and Wolfenden, R. (1977). Biochemistry 16, 4886-4890.

Lienhard, G.E. (1973). Science 180, 149-154.

Lofqvist, B. and Klevhag, J.-E. (1974). Acta Chem. Scand. B28, 1003-1012.

Lofqvist, B. and Sjoberg, L.-B. (1971). Acta Chem. Scand. 25, 1663-1678.

Lowe, G. and Nurse, D. (1977). J. Chem. Soc. Chem. Commun., 815-817.

Markland, F.S. and Smith, E.L. (1971). 'In' The Enzymes (Boyer, P.D., ed.), vol. 3, third edition, pp. 561-608, Academic Press, New York.

Markley, J.L. and Ibanez, I.B. (1978). Biochemistry 17, 4627-4640.

Matthews, B.W. (1968). J. Mol. Biol. 33, 491-497.

Matthews, B.W., Sigler, P.B., Henderson, R. and Blow, D.M. (1967). Nature 214, 652-656.

Matthews, D.A., Alden, R.A., Birktoft, J.J., Freer, S.T. and

Kraut, J. (1977). J. Biol. Chem. 252, 8875-8883.

McLachlan, A.D. and Shotton, D.M. (1971). Nature 229, 202-205.

Narahashi, Y. (1972). J. Biochem. (Tokyo) 71, 1077-1080.

Narahashi, Y. and Yoda, K. (1973). J. Biochem. (Tokyo) 73, 831-841.

Narahashi, Y., Shibuya, K. and Yanagita, M. (1968). J. Biochem. (Tokyo) 64, 427-437.

Narahashi, Y. and Yoda, K. (1977). J. Biochem (Tokyo) 81, 587-597.

Nomoto, M. and Narahashi, Y. (1959a). J. Biochem. (Tokyo) 46, 653-667.

Nomoto, M. and Narahashi, Y. (1959b). J. Biochem. (Tokyo) 46, 1481-1487.

Nomoto, M., Narahashi, Y. and Murakami, M. (1960a). J. Biochem. (Tokyo) 48, 593-602.

Nomoto, M., Narahashi, Y. and Murakami, M. (1960b). J. Biochem. (Tokyo) 48, 906-918.

North, A.C.T., Phillips, D.C. and Mathews, F.S. (1968). Acta Crystallogr. A24, 351-359.

Olson, M.O.J., Nagabhushan, N., Dzwiniel, M., Smillie, L.B. and Whitaker, D.R. (1970). Nature 228, 438-442.

Paterson, G.M. and Whitaker, D.R. (1969). Can. J. Biochem. 47, 317-321.

Pauling, L. (1948). Am. Sci. 36, 51-56.

Polgar, L. and Bender, M.L. (1969). Proc. Natl. Acad. Sci. (USA) 64, 1335-1342.

Poulos, T.L., Alden, R.A., Freer, S.T., Birktoft, J.J. and Kraut, J. (1976). J. Biol. Chem. 251, 1097-1103.

Powers, J.C. (1977). Methods Enzymol. 46, 197-208.

Ramachandran, G.N. and Mitra, A.K. (1976). J. Mol. Biol. 107, 85-92.

Ramakrishnan, C. and Ramachandran, G.N. (1965). Biophys. J. 5, 909-933.

Richards, F.M. (1968). J. Mol. Biol. 37, 225-230.

Richardson, J.S. (1976). Proc. Natl. Acad. Sci. (USA) 73, 2619-2623.

Robertus, J.D., Alden, R.A., Birktoft, J.J., Kraut, J., Powers, J.C. and Wilcox, P.E. (1972a). Biochemistry 11, 2439-2449.

Robertus, J.D., Kraut, J., Alden, R.A. and Birktoft, J.J. (1972b). Biochemistry, 11, 4293-4303.

Robillard, G. and Shulman, R.G. (1972). J. Mol. Biol. 71, 507-511.

Robillard, G. and Shulman, R.G. (1974a). J. Mol. Biol. 86, 519-540.

Robillard, G. and Shulman, R.G. (1974b). J. Mol. Biol. 86, 541-558.

Rossmann, M.G. and Argos, P. (1975). J. Biol. Chem. 250, 7525-7532.

Ruhlmann, A., Kukla, D., Schwager, P., Bartels, K. and Huber, R. (1973). J. Mol. Biol. 77, 417-436.

Sampson, P. (1970). 'In' Biomedical Computer Programs X-series Supplement (Dixon, W.J., ed.), pp. 177-186,

University of California Press, Los Angeles.

Sawyer, L., Shotton, D.M., Campbell, J.W., Wendell, P.L., Muirhead, H., Watson, H.C., Diamond, R. and Ladner, R.C. (1978). J. Mol. Biol. 118, 137-208.

Schechter, I. and Berger, A. (1967). Biochem. Biophys. Res. Commun. 27, 157-162.

Schoellmann, G. and Shaw, E. (1963). Biochemistry 2, 252-255.

Schultz, R.M. and Cheerva, A.C. (1975). FEBS Lett. 50, 47-49.

Segal, D.M., Cohen, G.H., Davies, D.R., Powers, J.C. and Wilcox, P.E. (1972). Cold Spring Harbor Symp. Quant. Biol. 36, 85-90.

Segal, D.M., Powers, J.C., Cohen, G.H., Davies, D.R. and Wilcox, P.E. (1971). Biochemistry 10, 3728-3737.

Shaw, E. (1970). 'In' The Enzymes (Boyer, P.D., ed.), vol. 1, third edition, pp. 91-146, Academic Press, New York.

Shotton, D.M. and Watson, H.C. (1970). Nature 225, 811-816.

Siegel, S. and Awad, W.M. (1973). J. Biol. Chem. 248, 3233-3240.

Sigler, P.B., Blow, D.M., Matthews, B.W. and Henderson, R. (1968). J. Mol. Biol. 35, 143-164.

Smillie, L.B. and Whitaker, D.R. (1967). J. Amer. Chem. Soc. 89, 3350-3352.

Stanier, R.Y., Doudoroff, M. and Adelberg, E.A. (1970). 'In' The Microbial World, third edition, pp. 147-158, Prentice-Hall Inc., New Jersey.

Steitz, T.A., Henderson, R. and Blow, D.M. (1969). J. Mol. Biol. 46, 337-348.

Stroud, R.M., Kay, L.M. and Dickerson, R.E. (1974). J. Mol. Biol. 83, 185-208.

Sweet, R.M., Wright, H.T., Janin, J., Chothia, C.H. and Blow, D.M. (1974). Biochemistry 13, 4212-4228.

Tatsuta, K., Mikami, N., Fujimoto, K., Umezawa, S., Umezawa, H. and Aoyagi, T. (1973). J. Antibiot. 26, 625-646.

Thiessen, W.E. and Levy, H.A. (1973). J. Appl. Crystallogr. 6, 309-346.

Thompson, R.C. (1973). Biochemistry 12, 47-51.

Trop, M. and Birk, Y. (1970). Biochem J. 116, 19-25.

Tsai, C.S., Whitaker, D.R. and Jurasek, L. (1965). Can. J. Biochem. 43, 1971-1983.

Tulinsky, A., Mani, N.V., Morimoto, C.N. and Vandlen, R.L. (1973). Acta Crystallogr. B29, 1309-1322.

Umezawa, H. (1976). Methods Enzymol. 45, 678-695.

Umezawa, H., Aoyagi, T., Morishima, H., Kunimoto, S., Matsuzaki, M., Hamada, M. and Takeuchi, T. (1970). J. Antibiot. 23, 425-427.

Venkatachalam, C.M. (1968). Biopolymers 6, 1425-1436.

Vosbeck, K.D., Chow, K.-F. and Awad, W.M. (1973). J. Biol. Chem. 248, 6029-6034.

Wahlby, S. (1969). Biochem. Biophys. Acta 185, 178-185.

Wahlby, S. and Engstrom, L. (1968). Biochem. Biophys. Acta 151, 402-408.

Walsh, K.A., Kauffman, D.L., Sampath Kumar, K.S.V. and

Neurath, H. (1964). Proc. Natl. Acad. Sci. (USA) 51, 301-308.

Weiner, H., White, W.N., Hoare, D.G. and Koshland, D.E. Jr. (1966). J. Amer. Chem. Soc. 88, 3851-3859.

Westerik, J. and Wolfenden, R. (1972). J. Biol. Chem. 247, 8195-8197.

Whitaker, D.R. (1965). Can. J. Biochem. 43, 1935-1954.

Whitaker, D.R. (1967). Can. J. Biochem. 45, 991-993.

Whitaker, D.R. and Roy, C. (1967). Can. J. Biochem. 45, 911-916.

Whitaker, D.R., Jurasek, L. and Roy, C. (1966). Biochem. Biophys. Res. Commun. 24, 173-178.

Whitaker, D.R., Cook, F.D. and Gillespie, D.C. (1965a). Can. J. Biochem. 43, 1927-1933.

Whitaker, D.R., Roy, C., Tsai, C.S. and Jurasek, L. (1965b). Can. J. Biochem. 43, 1961-1970.

Wilson, A.J.C. (1942). Nature, 150, 151-152.

Wolfenden, R. (1972). Acc. Chem. Res. 5, 10-18.

Wolfenden, R. (1976). Ann. Review Biophys. Bioeng. 5, 271-306.

Wright, H.T. (1973). J. Mol. Biol. 79, 1-11.

Wright, C.S., Alden, R.A. and Kraut, J. (1969). Nature 221, 235-242.

Wyckoff, H.W., Tsernoglou, D., Hanson, A.W., Knox, J.R., Lee, B. and Richards, F.M. (1970). J. Biol. Chem. 245, 305-328.

Zeppezauer, M., Eklund, H. and Zeppezauer, E.S. (1968). Arch. Biochem. Biophys. 126, 564-573.

Appendix 1

Amino Acid Designations

Amino Acid	Three-letter symbol	One-letter symbol
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic Acid	Asp	D
Cysteine	Cys	C
Glutamic Acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Reference: Biochem. J. (1969) 113, 1-4.

Appendix 2

Atomic Coordinates For The SGPA/Peptide Aldehyde Complex

Residue		Atom	X	Y	Z
PHE	P1	O	-14.1	20.1	27.0
PHE	P1	C	-14.0	20.8	25.5
PHE	P1	CA	-15.2	20.9	24.9
PHE	P1	CB	-16.0	19.7	24.3
PHE	P1	CG	-17.4	19.9	24.1
PHE	P1	CD1	-18.0	19.3	23.0
PHE	P1	CE1	-19.4	19.4	22.8
PHE	P1	CZ	-20.1	20.2	23.7
PHE	P1	CE2	-19.6	20.7	24.8
PHE	P1	CD2	-18.1	20.6	25.0
PHE	P1	N	-15.0	21.9	23.8
PRO	P2	O	-14.6	23.6	25.2
PRO	P2	C	-14.7	23.1	24.0
PRO	P2	CA	-14.5	24.1	22.8
PRO	P2	CB	-13.2	24.8	23.0
PRO	P2	CG	-13.4	26.3	22.9
PRO	P2	CD	-14.9	26.4	22.9
PRO	P2	N	-15.5	25.1	23.0
ALA	P3	O	-17.3	23.7	22.9
ALA	P3	C	-16.8	24.8	23.1
ALA	P3	CA	-17.8	26.0	23.3
ALA	P3	CB	-18.3	25.8	24.7
ALA	P3	N	-18.8	26.0	22.3
PRO	P4	O	-19.3	28.2	22.6
PRO	P4	C	-19.5	27.1	22.0
PRO	P4	CA	-20.6	26.9	21.0
PRO	P4	CB	-20.0	26.7	19.6
PRO	P4	CG	-19.9	28.0	18.9
PRO	P4	CD	-20.8	28.9	19.7
PRO	P4	N	-21.3	28.2	20.9
AC	P5	O	-22.4	28.0	22.8
AC	P5	C	-22.1	28.7	21.8
AC	P5	CA	-22.7	30.0	21.6
HIS	57	CB	-8.9	25.5	24.8
HIS	57	CG	-9.9	25.7	25.9
HIS	57	ND1	-10.6	24.6	26.5
HIS	57	CE1	-11.4	25.1	27.4
HIS	57	NE2	-11.3	26.4	27.4
HIS	57	CD2	-10.4	26.8	26.5
SER	195	CA	-12.3	17.8	24.4

(Appendix 2 continued)

SER	195	CB	-12.3	19.2	24.8
SER	195	OG	-13.6	19.8	24.4
WAT	1		-12.8	20.4	29.1
WAT	2		-14.4	21.4	30.8
WAT	3		-11.0	21.9	23.0
WAT	4		-10.5	21.9	26.5

Appendix 3

Atomic Coordinates For The SGPA - Chymostatin Complex

Residue		Atom	X	Y	Z
PHE	P1	O	-14.1	19.5	26.4
PHE	P1	C	-14.0	20.4	25.3
PHE	P1	CA	-15.4	20.9	24.7
PHE	P1	CB	-15.9	19.8	23.7
PHE	P1	CG	-17.4	19.9	23.5
PHE	P1	CD1	-18.2	20.3	24.6
PHE	P1	CE1	-19.6	20.4	24.4
PHE	P1	CZ	-20.2	19.9	23.3
PHE	P1	CE2	-19.4	19.5	22.2
PHE	P1	CD2	-18.0	19.5	22.3
PHE	P1	N	-15.0	22.1	24.2
LEU	P2	O	-16.2	23.1	25.9
LEU	P2	C	-15.5	23.1	24.8
LEU	P2	CA	-15.1	24.4	24.2
LEU	P2	CB	-13.9	25.1	24.8
LEU	P2	CG	-13.2	26.2	23.9
LEU	P2	CD1	-12.0	26.8	24.6
LEU	P2	CD2	-12.8	25.5	22.6
LEU	P2	N	-16.3	25.4	24.4
ARG	P3	O	-17.4	24.5	22.6
ARG	P3	C	-17.3	25.3	23.6
ARG	P3	CA	-18.4	26.3	23.9
ARG	P3	CB	-19.4	25.8	24.9
ARG	P3	CG	-19.9	24.4	24.9
ARG	P3	CD	-21.0	24.0	25.8
ARG	P3	NE	-22.1	25.1	25.7
ARG	P3	CZ	-21.8	26.3	25.3
ARG	P3	NEH1	-20.6	26.7	24.9
ARG	P3	NEH2	-22.8	27.2	25.4
ARG	P3	N	-19.1	26.1	22.5
URE	PX	O	-19.8	28.4	22.7
URE	PX	C	-19.8	27.3	22.1
PHE	P4	N	-20.5	27.2	20.9
PHE	P4	CD2	-22.2	27.7	17.3
PHE	P4	CE2	-22.3	26.6	16.3
PHE	P4	CZ	-21.3	25.7	16.3
PHE	P4	CE1	-20.1	25.8	17.1
PHE	P4	CD1	-20.1	26.8	18.0
PHE	P4	CG	-21.1	27.8	18.1
PHE	P4	CB	-21.0	28.9	19.1
PHE	P4	CA	-21.4	28.5	20.6
PHE	P4	C	-22.9	28.3	20.8

(Appendix 3 continued)

PHE	P4	O	-23.3	27.3	21.4
PHE	P4	O	-23.6	29.2	20.3
SER	195	CA	-12.3	17.7	24.4
SER	195	CB	-12.3	19.2	24.8
SER	195	OG	-13.3	19.9	24.1

Appendix 4

Atomic Coordinates For The SGPB - GLF Inhibitor Complex

Residue		Atom	X	Y	Z
PHE	P1	CH2	17.4	41.1	45.1
PHE	P1	O	17.8	43.2	44.3
PHE	P1	C	18.6	42.1	44.8
PHE	P1	CA	19.3	42.4	45.8
PHE	P1	CB	20.3	43.3	45.2
PHE	P1	CG	21.0	44.3	46.1
PHE	P1	CD1	22.4	44.4	46.1
PHE	P1	CE1	23.0	45.3	47.0
PHE	P1	CZ	22.2	46.0	47.9
PHE	P1	CE2	20.8	45.9	47.9
PHE	P1	CD2	20.2	45.0	47.0
PHE	P1	N	19.9	41.3	46.6
LEU	P2	O	18.8	41.8	48.5
LEU	P2	C	19.6	41.1	47.9
LEU	P2	CA	20.3	39.9	48.6
LEU	P2	CB	19.4	38.9	49.2
LEU	P2	CG	19.9	37.5	49.3
LEU	P2	CD1	18.9	36.6	50.0
LEU	P2	CD2	20.0	37.0	47.8
LEU	P2	N	21.1	40.6	49.6
GLY	P3	O	22.5	41.5	48.1
GLY	P3	C	22.1	41.3	49.3
GLY	P3	CA	22.9	41.9	50.5
GLY	P3	N	24.3	42.1	50.8
BOC	P4	O	24.0	41.7	53.0
BOC	P4	C	24.7	41.9	52.1
BOC	P4	OA	26.2	42.1	52.3
BOC	P4	CB	26.6	40.6	52.5
BOC	P4	CH1	25.5	39.7	51.9
BOC	P4	CH2	27.9	40.3	52.0
BOC	P4	CH3	26.5	40.5	54.1
HIS	57	O	13.6	37.8	44.8
HIS	57	C	14.1	36.9	44.1
HIS	57	CA	15.0	35.8	44.7
HIS	57	CB	16.1	36.1	45.8
HIS	57	CG	16.9	37.5	45.6
HIS	57	ND1	18.2	37.6	45.3
HIS	57	CE1	18.5	38.8	45.0
HIS	57	NE2	17.4	39.5	45.1
HIS	57	CD2	16.3	38.7	45.5
HIS	57	N	15.5	35.0	43.5

(Appendix 4 continued)

SER 195	O	17.5	42.6	39.4
SER 195	C	18.6	42.1	39.7
SER 195	CA	19.3	42.3	41.1
SER 195	CB	18.6	41.9	42.4
SER 195	OG	19.2	41.8	43.6
SER 195	N	19.2	43.8	41.1
TYR 171	O	22.0	35.2	56.9
TYR 171	C	21.6	35.2	55.7
TYR 171	CA	22.4	36.1	54.7
TYR 171	CB	21.6	36.7	53.6
TYR 171	CG	22.1	36.7	52.3
TYR 171	CD1	22.3	37.8	51.6
TYR 171	CE1	22.9	37.8	50.3
TYR 171	CZ	23.2	36.6	49.6
TYR 171	OEH	23.8	36.6	48.4
TYR 171	CE2	22.9	35.4	50.4
TYR 171	CD2	22.4	35.5	51.6
TYR 171	N	23.2	35.0	53.9

Appendix 5

Atomic Coordinates For The SGPB - AGF Inhibitor Complex

Residue		Atom	X	Y	Z
PHE	P1	CH2	17.3	41.1	45.2
PHE	P1	O	17.9	43.1	44.5
PHE	P1	C	18.5	41.9	44.8
PHE	P1	CA	19.4	42.1	45.9
PHE	P1	CB	20.4	43.2	45.5
PHE	P1	CG	21.0	43.9	46.6
PHE	P1	CD1	22.4	44.1	46.6
PHE	P1	CE1	23.0	44.8	47.7
PHE	P1	CZ	22.2	45.4	48.7
PHE	P1	CE2	20.8	45.2	48.7
PHE	P1	CD2	20.3	44.5	47.6
PHE	P1	N	20.1	40.9	46.5
GLY	P2	O	19.3	41.4	48.6
GLY	P2	C	19.9	40.7	47.8
GLY	P2	CA	20.7	39.4	48.3
GLY	P2	N	21.3	40.1	49.6
ALA	P3	O	22.9	41.2	48.4
ALA	P3	C	22.3	40.9	49.5
ALA	P3	CA	22.8	41.5	50.8
ALA	P3	CB	22.0	42.7	51.2
ALA	P3	N	24.2	41.8	50.7
BOC	P4	O	24.7	40.8	52.7
BOC	P4	C	25.1	41.4	51.6
BOC	P4	OA	26.5	41.8	51.4
BOC	P4	CB	27.3	40.8	52.2
BOC	P4	CH1	26.5	39.4	51.9
BOC	P4	CH2	28.7	40.6	51.7
BOC	P4	CH3	27.2	41.0	53.7
HIS	57	O	13.6	37.8	44.8
HIS	57	C	14.1	36.9	44.1
HIS	57	CA	15.0	35.8	44.7
HIS	57	CB	16.2	36.0	45.7
HIS	57	CG	16.9	37.4	45.6
HIS	57	ND1	18.2	37.6	45.3
HIS	57	CE1	18.4	38.9	45.0
HIS	57	NE2	17.3	39.5	45.2
HIS	57	CD2	16.3	38.6	45.5
HIS	57	N	15.5	35.0	43.5
SER	195	O	17.5	42.6	39.4
SER	195	C	18.6	42.1	39.7

(Appendix 5 continued)

SER 195	CA	19.3	42.3	41.1
SER 195	CB	18.6	41.9	42.5
SER 195	OG	19.1	41.4	43.7
SER 195	N	19.2	43.8	41.1